مبادئ الإحصاء

للاقتصاد والعلوم الإدارية

الأستاذ الدكتور
محمد الفاتح محمود بشير المغربي



مبادئ الإحصاء

للاقتصاد والعلوم الإدارية

بِسْمِ الـلـه الرَّحْمَنِ الرَّحِيمِ

مبادئ الإحصاء

للاقتصاد والعلوم الإدارية

الأستاذ الدكتور

محمد الفاتح محمود بشير المغربي

أستاذ إدارة الأعمال وعميد كلية الاقتصاد والعلوم الاجتماعية

جامعة القرآن الكريم والعلوم الإسلامية – السودان

المدرب المعتمد بالمجلس العام للبنوك والمؤسسات المالية الإسلامية

البحرين - المنامة

2022م



الأكاديمية الحديثة للكتاب الجامعي

الكتاب : **مبادئ الإحصاء للاقتصاد والإدارة**

المؤلف : أ. د. محمد الفاتح محمود بشير المغربي

رقم الطبعة : الأولى

تاريخ الإصدار : 2022م

حقوق الطبع : محفوظة للناشر

الناشر : الأكاديمية الحديثة للكتاب الجامعي

العنوان : 82 شارع وادي النيل المهندسين ، القاهرة ، مصر

تلفاكس : 561 33034 (00202) 0122/1734593

البريد الإليكتروني: m.academyfub@yahoo.com

رقم الإيداع : 10919 / 2021

الترقيم الدولي : 8 – 026 – 831 – 977 – 978

المقدمة

بسم اللـه والحمد لله والصلاة والسلام على خير البشر سيدنا محمد عليه الصلاة والسلام وعلى آله وصحبه أجمعين أما بعد...

يعتبر علم الإحصاء في الوقت الحالي واحد من أهم العلوم الحديثة التي تلعب دورا حيويا في كثير من العلوم والدراسات المختلفة. كما يعتبر الإحصاء من أقدم العلوم حيث ظهر مع حاجة الإنسان الأولى للتعامل مع القيم والأعداد التسيير الحياة اليومية. فالتاجر يسعى إلى حصر وحفظ البيانات المتعلقة بتجارته والمزارع يقوم دوما بإحصاء الإنتاج والمعلومات الأخرى المتعلقة كعدد الأشجار وأوقات الحصاد والبذر وغيرها من المعلومات والبيانات ذات العلاقة.

ومع التطور الهائل في العلوم كافة في أواخر القرن العشرين تطور علم الإحصاء ليستفيد من تقنيات الحاسب الآلي بشكل يجعله العلم الأكثر تداخلا مع العلوم الأخرى المختلفة، حيث أصبح يستخدم علم الإحصاء في العلوم التجارية وعلوم الطب والهندسة والأدب وجميع العلوم الأخرى دون استثناء. كما ساهم عصر المعلومات والانفتاح العالمي الحديث في إبراز أهمية تفعيل عملية التعامل مع البيانات بأسلوب يضمن السيطرة عليها وقراءتها، مما كان له الأثر الواضح على تطور علم الإحصاء كونه العلم الذي يحقق تلك الغاية. كما اتجهت كثير من العلوم والدراسات الأكاديمية والبحثية لاسيما التطبيقية إلى استخدام علم الإحصاء من خلال حصر بيانات مشكلة البحث والتعامل معها إحصائيا للوصول إلى فهم أفضل وحلول موضوعية.

يتم الاستفادة من علم الإحصاء في مجالات متنوعة تشمل ميادين عديدة كالصناعة والزراعة والطب والبحوث وغيرها من مجالات الإدارة والأعمال والاقتصاد والعلم بشكل عام. ويتم تطبيق الأساليب الإحصائية في الجوانب المختلفة للصناعة كمراقبة جودة المنتجات وتسويقها والتخزين وتشغيل خطوط الإنتاج. كما يتم دراسة السكان والمساكن من خلال الإحصاء الديموجرافي، حيث يتم التركيز على القوى العاملة وخصائصها والأجور والدخل والإنفاق. أما في مجال الأعمال والتجارة فان الإحصاء يلعب دورا

حيويا يتمثل في دراسة السوق واتجاهات المستهلكين ودراسات الأسعار وكميات الإنتاج.

وبهذا يكون هذا الكتـاب بـإذن اللــه مناسـباً للأشـخاص الـراغبين تنميـة مهـاراتهم الإحصائية في مجال الادارة والتجارة والمجالات الاخرى.

والـلـه خير الموفقين وعليه توكلنا،،،

المؤلف

الفصل الاول

تطور علم الإحصاء

يتعامل الناس عموماً في حياتهم اليومية مع المفاهيم وحتى بعض المفردات الإحصائية وبالأخص ما يتعلق منها بالاحتمالات وبعض المقاييس الوصفية مع رصد ما يطرأ عليها من تغيرات عبر فترات زمنية متعاقبة. وإذا كان هذا الأمر محسوساً بالنسبة لنا في الوقت الحاضر، فإن حقب التأريخ القديم وما بعده أفرزت أحداثاً تنم مجرياتها عن بدء استخدام الأساليب الإحصائية على أرض الواقع التي يمكن اعتبارها أفكاراً تتناغم مع بعض أحداث الأساليب الإحصائية المعاصرة.

وبالرغم مما نشهده من تطور متسارع في علم الإحصاء ليشمل كافة جوانب العمليات والطرق الإحصائية، إلا أن انعكاسات ذلك في العمل الإحصائي العربي لم يكن بالمستوى الذي يمكن تلمسه من قبل العديد من العاملين في المجال الأكاديمي وأولئك العاملين في الميدان الإحصائي ضمن المؤسسات العامة أو الخاصة.

وعلى عكس ما هو ملموس في الدول المتقدمة التي ظهرت فيها حب متتالية من التطوير لجوانب مختلفة من علم الإحصاء نتيجة للتفاعل والتعاون القائم بين العلميين الذين ساهموا في هذا التطوير والمؤسسات التي كانت بحاجة له، نجد أن الأمر لدينا حالياً متمركز على طروحات من الطرفين يمكن وضعها بشكل شكاوى أو إتهامات متبادلة من أحد الطرفين للطرف الآخر. ففي الوقت الذي يحاول فيه الاحصاء الاكاديمي لدنا متابعة المستجدات في البرامج الإحصائية التطبيقية في غالبية الجامعات العالمية وتبني ما هو مناسب لتحديث البرامج الدراسية طبقاً لذك، يشكو فيه العاملون في الميدان الإحصائي ابتعاد هذه البرامج عن الغايات التطبيقية للاساليب الإحصائية وتركيزها على الجانب النظري ذو الطابع الرياض (من خلال قراءتهم لمساقات متعددة في النظريات الإحصائية والرياضيات) ما يخلق إشكاليات وصعوبات لدى الخريجين بتخصصات إحصائية عند عملهم في الميدان.

من جانب آخر، يحاول الطرف الاكاديمي التوضيح بأن آلية التحديث المستمر للبرامج الدراسية تأخذ في الاعتبار مجالات التطبيق للأساليب الإحصائية في عمل المؤسسات العامة والخاصة ولا يوجد من المسافات النظرية في الإحصاء والرياضيات غلا ما هو بحكم الضرورة لتمكين الطالب من استيعاب المنطق الحسابي والتحليلي للبيانات الإحصائية عند التطبيق.

وفي هذه الورقة نحاول تناول عدد من الجوانب في العمل الإحصائي للمساعدة في توضيح الصورة ووضع الأمور في نصابها قد الأمكان لمد المزيد من جسور التفاعل والعمل المشترك بين الفريقين الاكاديمي والميداني. من خلال الطواف في مجربات التطور التاريخي لعلم الاحصاء، سيتم إلقاء الضوء على العلاقة المتبادلة مابين النظرية والتطبيق في علم الإحصاء والتي يمكن وضعها ضمن مفهوم التغذية المتبادلة لكل منهما للآخر مما دفع علم الإحصاء إلى أن يشهد التطور الكبير الذي طرأ عليه وهو ما نلمسه حالياً.

البعد التاريخي في علم الإحصاء:

عند الكلام عن علم الأحصاء وتطوره تاريخياً، غالباً ما يكون في ذهننا الذهاب إلى بدايات القرن السابع عشر وربما أحياناً القرن السادس عشر على أبعد تقدير. والدافع لذلك بطبيعة الحال هو ما تعرفه عن بدايات العمل في امور حياتية والتعامل مع معطياتها بصيغ يغلب عليها الربط مع المنطق الرياضي السائد آنذاك. ولكننا على أية حال يجب أن لا يغيب عن بالبنا ماورد في القرآن الكريم من ذكر لكلمة الإحصاء كدلالة لفكرة العد والحصر وهو أقدم من ذلك بقرون عدة. وجدير بالذكر ان ثمة ممارسات تطبيقية قد حدثت في التأريخ القديم الذي يمتد إلى زمن النبي نوح (عليه السلام) وأن استيعابنا لسمة المنطق الذي كان يحكمها يدفعنا لوضعها ضمن العمل الإحصائي بل واعتبارها أساساً لطرق إحصائية معرفية تم تطويرها واستخدامها في التطبيقات الإحصائية الحديثة.

ويذكر أنه بعد مرور أربعين يوماً على الطوفان، أراد النبي نوح (عليه السلام) ان يستطلع الأمر فأرسل الغراب من على السفينة إلا أنه ظل يذهب ويجئ دون أن يستنتج منه النبي نوح (عليه السلام) أي شئ فيما يتعلق بما آل

إليه الطوفان وهو معرفة ما إذا بدأ الماء بالانحسار وظهور اليابسة. بعد ذلك أرسل الحمامة على فترات زمنية متعاقبة انتهت بمجيء الحمامة في المرة الأخيرة وهي تحمل في منقارها غصن الزيتون لعلها تبني به عشاً على السفينة. عندها استنتج النبي نوح (عليه السلام) بن انحساراً للماء وظهوراً لليابسة قد بدأ وأن السلامة لمن هم على ظهر السفينة قد تحققت وهذا ما أوحي للبعض أن يستخدم شعار الحمامة مع غصن الزيتون رمزاً لسلام كما هو معروف.

فإذا اعتبرنا الفترات الزمنية المتعاقبة بمثابة مستويات الجرعة Dose Levels ودليل ظهور اليابسة بالاستجابة النوعية Quantal Response، لا يمكننا إلا اعتبار هذا أساساً لأسلوب (الاستجابة النوعية في التجارب الحيوية (Quantal Response Tecnique in Bioassay) والذي هو أحد أحدث الساليب الإحصائية في الوقت الحاضر والأكثر تطوراً واستخداماً.

وفي صدر الاسلام يعتبر الاسلوب الذي كان الخليفة عمر بن الخطاب "رضي الله عنه" يستخدمه لتقدير عدد المقاتلين من خلال معلومة عن عدد أرغفة الخبز المستهلكة إلا أفكاراً تتناغم كلياً مع اسلوب (معلومات المتغير المساعد Auxiliary Variable Information) والمستخدم حالياً في اساليب التقدير لمتغيرات يصعب أخذ معلومات عنها في المعاينة. هذا بالاضافة لإجراء احصاء عام وتدوين الدواوين في عصر خلافته (643-634).

عصر الاحتمالات والإحصاء:

من المعروف أن ثمة بدايات معروفة في مجال الاحتمالات قد ظهرت في القرن السادس عشر حيث قد Cardano(1571-1501) بعض الافكار في الاحتمالات المرتبطة برمي زهرة الطاولة. وقعد ذلك تطور العمل في مجال الاحتمالات وظهرت الطرق الإحصائية بابعادها النظرية والتطبيقية، وتعتبر الرسائل والنقاشات التي تدور بين Pascal(1623-1662) وFermat مؤشراً لظهور أصول الاحتمالات حينما عالجا بعض المسائل المرتبطة بالعاب الحظ. وكان Fascal قد قدم عام 1665 أسس التوقع وناقش مسألة إفلاس المراهنين.

إلا أن البعد الإحصائي بمفهومه النظري (الرياضي) والمعروف حالياً قد شهد حقبة تاسيس وتطوير بدأت في القرن الثامن عشر وامتدت إلى الثلث الاول من القرن العشرين. وفي البدء، لم يكن التطوير في نظريات الاحتمال والطرق الإحصائية الا استجابة لحاجات تطبيقية حقيقية في العلوم وقضايا المجتمع. وقد اسهم Laplace(1749-1827) في ترسيخ مفهوم عمومية التطبيق للطرق الإحصائية بشكل عام وأثبت كون النظرية الاحتمالية اسلوباً ضرورياً لتحسين جميع أنواع المعرفة الانسانية. فقد أوضح امكانية التطبيقات في مجالات الالعاب المبنية على الحظ، العلوم الطبيعية مثل (علم الفلك، علوم الأرض، علم المناخ)، العلوم الانسانية مثل (صدقية الاستجواب والحكم، علم التشريع، الانتخابات، قرارات اللجان)، علوم السكان ان الإحصاءات الحياتية، التأمين على الحياة.

وبشكل عام، نرى ان الطرق الإحصائية التي كان يتم تطويرها آنذاك لتلائم العمل التحليلي في حقل ما من العلوم، تكون فيما بعد مناسبة للتطبيق في مجالات اخرى أو تطويرها باتجاه ما من قبل آخرين لتكون كذلك. فنجد فى سبيل المثال أن Quetelet (1796-1874) وهو عالم فلك وإحصائي تعلم شيئاً عن منطقية الاحتمالات خلال رحلة علمية غلى باريس عام 1824 وعمل على التطبيقات في العلوم الاجتماعية وطالب بادخال تحسينات على عملية التعداد باستخدام هذه التطبيقات. كذلك عمل كلا من Galton(1857-1936) و Pearon(1822-1911) بالنسبة للتطبيقات في حقلي الوراثة وعلوم الحياة، وما تم تطويره من قبل Fisher(1890-1962)في حقلي الجينات والتجارب الحقلية الزراعية يدخل في هذا الاطار. وأن عمل هؤلاء على تطبيق طرق إحصائية في المجالات المذكورة قادهم إلى تطوير احصائية جديدة.

هدت السنوات الأولى من فترة (1920-1955) والتي سبقت قيام الحرب العالمية الأولى (1914-1918) تطوراً كبيراً وامتداداً واسعاً للاحتمالات والإحصاء في كل الاتجاهات. إلا أن الحرب التي أثرت بشكل كبير في جميع النواحي كان تأثيرها واضحاً في العمل الإحصائي حيث توقف البحث تقريباً في هذا الجانب وذلك بسبب انخراط الناس في الفعاليات العسكرية

والقيام بأعمال أخرى تخص الجانب الحربي Pearson على سبيل المثال عمل في مجال القذائف، Jeffreys في مجال المناخ وyule في الإدارة.

ومن الملاحظ ظهور مساهمات في مواضيع أخرى وجدت في النهاية مكاناً لها ضمن نظرية العمليات التصادفية (العشوائية). ففي الفيزياء مثلاً عمل Einstein و Smoluchoveski على الحركة البراونية بينما Bacheljer طور نموذجاً مشابهاً لاستخدامه في التخمين المالي. كذلك طور Lundbberg وهو الخبير في شؤون التامين نظرية المخاطرة الجماعية. ونجد ايضاً أن الملاريا وهجرة البعوض كانت وراء اهتمام Pearsonفي موضوع الميسر العشوائي (Random Wolk) وكان هناك ايضاً نماذج رياضية في علم الأوبئة تم تطويرها من قبل Mckendrick and Ross.

ومع أن Mendel لم يستخدم الاحتمالات في عمله بموضوع الجينات (1866) لكن فكرته قد تم تطويعها فيما بعد احتمالياً من قبل Pearson, Fisher and Yule من خلال التحقق من المدى الذي يمكن لأسسه التي طرحها من التقارب مع النتائج التي يجدها علماء القياس (النماذج) في علم الحياة.

ومن خلال (1863-1945) أصبح موضوع الارتباط ذو أهمية واضحة في علم النفس بعد مساهماته الإحصائية مثل الارتباط الرتبي والتحليل العاملي والذي اضاف عليه Thurstone فيما بعد التحليل العاملي المتعدد عام 1930.

وخلال هذه الفترة، أصبحت طرق التحليل الكمي شائعة الاستخدام في حقل الاقتصاد في الولايات المتحدة، وأن ما قدمه كل من Moor Mitchell , person,i Lrving Fisher يمكن تصنيفه أساساً تحليلي السلاسل الزمنية. كما بدأ تطبيق الاحتمالات في المجال الصناعي مع عمل Erlag حول الاختناقات في أنظمة الهاتف والتي تعتبر أساساً لنظرية الطوابير حالياً.

ومن الجدير بالذكر أن هذه الفترة شهدت تقدماً ملحوظاً في مجال المؤسسات التعليمية الإحصائية تضمنت تأسيس قسم الإحصاء التطبيقي عام 1911 في جامعة كاليفورنيت/لوس أنجلوس ULC والذي ترأسه Pearson، كذلك أصبح Bowley أول من يسمى أستاذاً في الاحصاء في USE /لندن في بريطانيا عموماً. وفي جامعة كمبردج تم اتحدا مسمى أكاديمي بعنوان

(محاضر إحصائي) حيث كان (1971-1951) أول من حصل عليه والذي قد يمكن تسميته بأول إحصائي حديث وقد كان يهتم بتطبيق كل ماكان يسهم به في الاحصاء. ونتيجة اهتمامه بنظرية مندل قادة ذلك إلى إيجاد طريقة أصغر مربع كاي في التقدير. و(1856-1922) الذي عاصر تلك الفترة أسهم في إيجاد نظرية الحد المركزي وقانون الأعداد الكبيرة ومنثك قدم سلسلة ماركوف المعروفة. كما قام بتطوير نظرية Gauss التي قدمها عام (1821) وسميت بعد ذلك بنظرية ماركوف – غاوس.

ومن الذين شهدت تلك الفترة اسهاماتهم كان Gosset as (1876-1937) w.s Students الذي نشر أول بحث له تحت اسم (Student) حيث تضمنت إعادة اكتشاف توزيع بواسون. وفي عام (1908) طرح موضوع توزيعات العينات الصغيرة من خلال ورقتين تناولت إحداها طبيعة التوزيع لمتوسط العينة والذي أصبح معروفاً بـ (Student's distribution and Studentization) والأخرى حول الارتباط الطبيعي. كما أنه نادى بإمكانية استخدام طريقة فرق التغاير للتعامل مع الارتباط الوهمي.

فترة (1930-1940) شهدت هذه الفترة تطورات مهمة في الاحتمالات، النظرية الإحصائية والتطبيقات الإحصائية. ففي الاحتمالات، كانت التطورات الرئيسية تتمثل في بديهيات Kolmogorov(1903-1987) لاحتمال إضافة إلى تطويره للنظرية العامة للعمليات التصادفية في Khinvhin(1894-1959) وهذا العمل اعتبر مؤشراً لبداية الاحتمالات المعاصرة.

وفي بريطانيا والولايات المتحدة بدأت فترة إعادة تعريف للاحصاء. فالجمعية الإحصائية الملكية خرجت من محيط الإحصاء الرسمي وأصبحت ترحب بأعمال إحصائية في مجالات الزراعة والصناعة وكذلك الإحصاء الرياضي. وكان هناك تغيراً مماثلاً في الجمعية الإحصائية الأمريكية عندما توقفت مجلة Biometrika عن نشر بحوث حيوية وركزت على الإحصاء النظري. كذلك نجد أن معهد الإحصاء الرياضي قد تأسس عام (1930) وبدأ بإصدار مجلته The Annals of Mathematical Statistics عام (1933) والتي أصبحت المجلة الرئيسية في الإحصاء الرياضي والاحتمالات.

وفي الاستقصاء الإحصائي كانت التطورات الرئيسية متمثلة بنظرية Pearson – Neyman لاختبار الفرضية والتي بدأ العمل عليها منذ عام (1928م).

الإحصاء الحديث وشواخصه في المجال التطبيقي:

بالرغم من الإحساس السائد بأن الجانب النظري كان يطغى على ملامح التطوير لعلم الإحصاء الحديث بشكل عام والذي بدأ مع نهاية القرن التاسع عشر واستمر خلال القرن العشرين، فإن المشهد لهذا التطوير ينم عن أنشطة تطبيقية واسعة شكلت بحد ذاتها محطات شاخصة في العمل الإحصائي عبر تكلم الحقبة الزمنية. ويمكن النظر إلى هذه الانشطة التطبيقية الإحصائية من جانبين، ففي الوقت الذي قد نجد فيها ما يمكن اعتباره تطبيقاً مباشراً لطرق إحصائية معروفة في حينها ضمن مجالات اجتماعية وطبية واقتصادية وغير ذلك، نجد أيضاً ثمة تطوير لطرق إحصائية جديدة جاءت استجابة لمتطلبات التحليل الإحصائي لبيانات في مجالات متنوعة عايشها الإحصائيون من خلال عملهم ضمنها أو استجابتهم لحاجة العاملين فيه لمثل هذه المتطلبات.

بدايات تشكيل الإحصاء الأكاديمي:

- في عام 1911 تأسس الإحصاء التطبيقي من قبل Pearson في الكلية الجامعية في لندن.
- في عام 1931 ساهم Hoielling في تأسيس قسم الإحصاء في جامعة كولومبيا.
- في عام 1939 عمل Cochran على تأسيس برنامج الدراسات العليا في الإحصاء ضمن قسم الرياضيات في جامعة أيوا ISU.
- Leslie Kish(1910-2000) ساهم في تأسيس معهد بحوث المسوحات في جامعة آنه آرير عام 1941.
- عام 1947أسس Snedecor أول قسم للإحصاء بشكله المستقل في ISU بعد أن كانت درجات الماجستير وحتى الدكتوراة في الإحصاء تمنح من قسم الرياضيات. وجاء قسم الإحساء هذا تواصلاً للعمل الاستشاري الإحصائي في الجامعة المذكورة من خلال أول مركز للاستشارات الإحصائية في الجامعة والذي تأسس عام 1933م.

الفصل الثاني

علم الاحصاء ووصف البيانات

Statistics and Data Description

مقدمة وتعريف علم الإحصاء:

وردت كمة الإحصاء في عدة آيات كريمة في القرآن الكـريم في سـورة إبـراهيم، سـورة مريم، وفي سورة الكهـف. وهـذا دلالـة عـلى أن علـم الإحصـاء قـديم العهـد. كـما يعتـبر المصريين من الأوائل الذين استخدموا علم الإحصاء وطبقوهـا في بنـاء الأهرامـات وقـاموا بتعداد لسكان البلاد وثروتها واستخدموا النتائج في تنظيم مشروع البناء.

وكذلك في عصر الدولة الإسلامية تم استخدام العد في معرفـة عـدد السـكان ومقـدار الزكاة وكان استخدام الإحصاء قديماً مقصوراً على الاعمال الخاصة بشـؤون الدولـة حيـث أن كلمة (Statistics) الإحصاءات مشتقة من كلمـة الدولـة State، وتعنـي مجموعـة أو أكثر من البيانات العدديـة عـن السـكان والـثروة والتجـارة الخارجيـة والإنتـاج الصـناعي والزراعي والضرائب...الخ التي تهم الدولة.

ويعتبر العالم الإسلامي فيشر (R.A.Fisher) أشهر علماء القرن العشرين حيـث طـور علم الإحصاء وطبقه في علوم كثيرة مثل الزراعة، الاقتصاد، الوراثة... الخ.

ومعنى إحصاء لأي فرد فيقتصر على الجداول العددية التي تصف ظاهرة خلال فترة زمنية معينة، ويمكن ملاحظة ذلك من خـلال تصـفح الصـحف أو المجـلات حيـث يمكـن مشاهدة بعض الجداول التي تبـين ارتفـاع اسـعار الـنفط أو عـدد الحـوادث المروريـة أو تمثيل بياني عن عدد السكان.

تعريف علم الإحصاء:

يعرف علم الإحصاء بعلم العد، وهو العلم الذي يهتم بوصف طرق متعـددة لجمـع البيانات والمشاهدات ومن ثم تنظيمها وعرضـها باسـتخدام الأسـاليب العلميـة لتحليلهـا واستخلاص النتائج منها بهدف الوصول إلى اخذ قرارات مناسبة.

يعرف علم الإحصاء بأنه: ذلك الفرع من العلوم الذي يختص بالطرق العلمية لجمع البيانات وتنظيمها وتلخيصها وعرضها وتحليلها وذلك للوصول إلى نتائج موثوقة لدعم اتخاذ قرارات سليمة على ضوء هذا التحليل. وسوف نتناول بعون الله تعالى آل طريقة بالشرح المفصل والأمثلة التوضيحية.

وهناك علاقة وطيدة بين علم الإحصاء والعلوم الأخرى مثل الرياضيات، علم اجتماع، التعداد السكاني، العلوم الإنسانية، العلوم الطبية والهندسية وغيرها من العلوم.

ويمر علم الإحصاء بالمراحل الخمس التالية:

1. الجمع: يعتبر الأساس في التحليل الإحصائي فإذا كانت البيانات غير دقيقة فإن الاستنتاج والقرارات المبنية عليها تكون غير دقيقة. لذلك يجب الدقة في الحصول على البيانات سواء من المصادر المنشورة أو غير المنشورة أو يجمعها من الميدان.
2. التنظيم: إذا أخذت البيانات من المصادرة المنشورة أو غير المنشورة غالباً ما تكون منظمة، أما التي تحصل عليها من الميادين باستخدام المسوح الإحصائية فإنها بحاجة إلى تنظيم. ويتم ذلك من خلال معالجة لمشاكل والتباين في المعلومات وعدم علاقتها بموضوع البحث أو الدراسة، ومن ثم تصنيفها وتبويبها على شكل جداول تكرارية Frequency Tables(7-2).
3. التقديم: تقديم البيانات والمعلومات من خلال عرضها بأشكال هندسية أو رسومات بيانية، عرض البيانات (14-2).
4. التحليل: أساليب التحليل كثيرة ومتعددة تمتد من المشاهدة البسيطة إلى الأساليب الرياضية المعقدة. والتحليل الإحصائي يركز على البيانات المبوبة في الجداول التكرارية.
5. التفسير: تعتبر هذه المرحلة من اهم مراحل البحث الإحصائي وتحتاج إلى درجة عالية من المهارة والخبرة وذلك بسبب استخلاص النتائج من البيانات التي تم جمعها وتحليلها وفي ضوء ذلك يتم اتخاذ القرارات المناسبة.

اقسام علم الإحصاء:

قسم علماء الإحصاء إلى قسمين رئيسيين هما:

أولاً: الإحصاء الوصفي: Descriptive Statistics:

يشتمل على تمثيل الطرق الإحصائية في جمع البيانات والمعلومات لتلخيصها واختصارها ومن ثم عرض المعلومات عن طريق الجداول أو الرسوم البيانية. فالإحصاء الوصفي يهتم بطرق جمع البيانات وتمثيلها وعرضها

ثانياً: الإحصاء الاستدلالي Inferential Statistice

ويسمى بالإحصاء الاستنتاجي أو التحليلي لأنه يعني بتحليل البيانات المتوفرة في العينة وتفسير النتائج بهدف التوصل إلى الأساليب التقدير والاختبار واتخاذ القرارات والتنبؤ أو الاستقراء، والإحصاء الاستدلالي يهتم بتحليل وتفسير البيانات والتوصل إلى الاستنتاجات.

خطوات البحث العلمي:

يعتمد البحث العلمي على الطرق والأدوات الإحصائية المختلفة حتى يتم استخدامها في العلوم او في أي مجال آخر وحسب طبيعة ونوع البحث المرغوب.

ويمكن تلخيص اهم خطوات البحث العلمي بالمراحل التالية:

1. تحديد المشكلة: يتم تحديد نوع المشكلة التي تستحق البحث والدراسة والتقصي ويكون دور الباحث في كيفية اختيار المشكلة المناسبة لدراسة ظاهرة ما كما يجب ان توضح عملية الاختيار العلاقة بين المتغيرات التي تشملها الدراسة أو المشاهدات بحيث تمكن الباحث من إجراء التحليل الإحصائي أو الاستنتاجي.
2. تحديد الأساليب الإحصائية: بعد تحديد مشكلة الدراسة ومعرفة جوانبها والطرق التي سوف تستخدم لحل المشكلة وفي ضوء ذلك يستطيع الباحث من اختيار البيانات المناسبة حتى يستكمل خطوات الدراسة.
3. مرحلة جمع البيانات: تعتمد على بعض الأساليب الإحصائية في جمع المعلومات أو البيانات، وذلك من خلال المصادر المباشرة أو غير المباشرة.

4. تحليل البيانات: استخدام الطرق والأساليب الإحصائية المختلفة في تحليل البيانات المتوفرة في العينة.
5. استخلاص النتائج ووضع التوصيات: يتم استناداً على التحليل الإحصائي لبيانات العينة لوصف النتائج حول مجتمع الدراسة ومن ثم اقتراح الحلول لمشاكل ووضع التوصيات المختلفة.
6. تعميم النتائج: لابد من الاستفادة من الدراسة وهنا يقوم الباحث بتعميمي النتائج على مجتمع الهدف مثلاً: إذا قام باحث بدراسة المشاكل التي تواجه التعليم في المرحلة الابتدائية وعند التوصل إلى نتائج ووضع توصيات فمن الضروري تعميم النتائج على جميع المهتمين بقطاع التعليم.

تعاريف:

المجتمع: مجموعة العناصر أو الافراد أو الشاهدات التي ينصب عليها الاهتمام في البحث أو الدراسة.

العينة: هي جزء من المجتمع.

مجتمع العينة: هو المجتمع الذي نأخذ منه العينة.

مجتمع الهدف: هو كل المجتمع الذي نطلب المعلومات عنه وستعمم عليه نتائج الدراسة

مجتمع الدراسة: هو مجموعة الأفراد والمشاهدات التي يتاح لنا إجراء الدراسة عليها.

مصادر جمع البيانات:

غالباً ما تقسم مصادر جمع البيانات إلى قسمين وهما:

أ. التقارير الرسمية التي تنشرها المؤسسات والجهات المخولة بذلك.

ب. الأفراد والمؤسسات التي تقوم بجمع البيانات من ذوي العلاقة.

ويمكن تقسيم مصادر جمع البيانات إلى قسمين كما يلي:

أولاً: المصادر المباشرة (الأولية):

المصدر الأولي لجمع البيانات هو ذلك المصدر الذي يقوم بجمع البيانات بنفسه او تحت إشرافه. مثل ما تقوم به دائرة الإحصاءات العامة من تسجيل البيانات عن السكان، ومن طرق المصادر المباشرة للبيانات:

أ. المقابلات الشخصية المباشرة: حيث يقوم جامع البيانات بطرح الاسئلة ويجيب عليها الشخص المعني (وجهاً لوجه).

ب. المقابلات الشخصية غير المباشرة: يقوم جامع البيانات بمقابلة شخص ثالث غير الشخص المطلوب منه البيانات حيث أن هذا الشخص لا يرغب بإعطاء المعلومات أو البيانات عن نفسه أو يكون غير متوفر في فترة جمع البيانات.

ت. المعلومات عن طريق المراسلين: دورهم جمع البيانات وتقديمها للجهة المعنية بالدراسة.

ث. عن طريق الهاتف: وهذه الطريقة شبيهة بالمقابلة الشخصية ولكن تكون الاسئلة والإجابات عن بعد عن طريق الهاتف.

ج. الاستبانات بالبريد أو الفاكس أو البريد الالكتروني، حيث يرسل الباحث الاستيانة للافراد المطلوب جمع البيانات منهم ويتم اعادة الاستبانة بعد التعبئة بنفس الطريقة.

ح. عن طريق الانترنت حيث يتم طرح الموضوع قيد الدراسة على الانترنت فيجب عليه الراغبون في المشاركة.

ثانياً: المصادر غير المباشرة (الثانوية):

عندما لا يستطيع الباحث من جمع البيانات بنفسه أو تحت اشرافه يلجأ إلى المصادر غير المباشرة أي إلى البيانات التي جمعها غيره: وهي معدة مسبقاً من طرف الجهات لمعنية والتي غالباً ما تكون رسمية، وتقسم هذه البيانات إلى قسمين وهما:

أ. المصادر المنشورة ومنها:

- التقارير والمنشورات الرسمية وهي تقارير ومنشورات تنشرها هيئات محلية مثل تقارير البلديات وغرف التجارة والصناعة المحلية.
- التقارير والمنشورات الخاصة، مثل تقارير ومنشورات الشركات والمؤسسات الخاصة (بنك التنمية الإسلامي).

ب. المصادر غير المنشورة:

وهي بيانات غير منشورة لكنها مدونة في سجلات الهيئات ويكون في إمكان الباحث الرجوع إليها عند الحاجة مثل البيانات عن المواليد في مدينة الزرقاء حيث يقوم الباحث بزيارة المستشفيات ويسجل ما يريد من البيانات المسجلة في السجلات اليومية التي تحتفظ بها المستشفيات.

طرق جمع البيانات Data Collection

يحتاج الباحث إلى البيانات الضرورية من اجل إنهاء البحث أو الدراسة التي يرغب بها لذلك يتم جمع البيانات الإحصائية بإحدى الطرق التالية:

1) طريقة المسح الشامل Census Method

حيث يتم جمع البيانات الإحصائية عن جميع المفردات التي تؤلف المجتمع الإحصائي قيد الدراسة مثل التعداد السكاني للدولة أو التعرف على مستوى طلاب الجامعة في مادة الاقتصاد أو حصر إعداد الطلبة في الجامعة.

2) طريقة العينة Sample Method

حيث يتم جمع البيانات عن جزء من وحدات المجتمع الإحصائي وذلك في حالة تعذر إجراء المسح الشامل وعندها تلجأ إلى دراسة جزء من المجتمع الإحصائي يسمى العينة وحجمها هو عدد عناصرها.

تكون الطريقتان محل مفاضلة إذا أمكن تعريف مجتمع الدراسة بالكامل، وفي هذه الحالة فإن استخدام أحد الأسلوبين تحكمه بعض الاعتبارات مثل، طبيعة المجتمع وطبيعة البيانات المطلوبة والإمكانيات الفنية المتاحة والوقت اللازم لإجراء الدراسة إضافة إلى الدعم المادي.

ومن الأسباب التي تؤدي إلى استخدام العينات بدل من المسح الشامل:

- توفير الوقت والجهد والنفقات.
- إذا كان المجتمع الإحصائي متجانساً.
- إذا كان المجتمع الإحصائي غير محدود.
- فساد عنصر المجتمع نتيجة أخذ المشاهدات، فإذا أرادت دائرة الصحة دراسة صلاحية معلبات وصنع ماء في هذه الحالة لا تستطيع أخذ جميع المعلبات التي يتم تصنيعها فإنها تختار عينة فقط.

- قد يكون المجتمع متصلاً وغير قابلة للعد، مثل دراسة المخزون من النفط.

العينات وطرق اختيارها Samples

العينة عبارة عن جزء من مجتمع الدراسة، وهي عبارة عن مجموعة الخطوات أو الإجراءات لاختيار هذا الجزء من أجل الحصول على استنتاجات متعلقة بمجتمع الدراسة، ومن خلالها يستطيع الباحث الحصول على فكرة مبدئية أو انطباع أولي عن بعض الأمور المتعلقة بموضوع الدراسة.

وإمكانية التعميم من العينة إلى المجتمع تعتمد على كيفية أخذ العينة وحجمها وطرق دراسة صفاتها باستعمال نظرية الاحتمال.

طرق اختيار العينة Sampling Techniques

الغاية هنا اختيار عينة تمثل المجتمع وتؤدي إلى إحراز معلومات عن سمة المجتمع بشكل دقيق يتناسب مع التكلفة والجهد المتعلمين وتقسم العينات بشكل عام إلى نوعان هما:

أولاً: العينات الاحتمالية Probability Sampling

يرتبط اختيار أية وحدة من وحدات المجتمع باحتمال معين ويمكن قياس خطأ المعاينة العشوائي ومنها:

1) العينة العشوائية البسيطة: Simple Random Sample

تستخدم العينة العشوائية البسيطة في الحالات التي يمكن فيها افتراض تجانس مجتمع الدراسة. ولها طريقتين بـ:

الطريقة الأولى: عندما يكون حجم المجتمع صغيراً

مثال: فإذا فرضنا أن هناك مجتمعاً حجمه N-10 وأردنا اختيار عينة عشوائية بسيطة حجمها n-5. تقوم عملية الاختيار على ترقيم المفردات على بطاقات متماثلة تماماً على كل بطاقة رقماً من 1 حتى 10 ومن ثم وضع البطاقات في كيس وتخلط جيداً ثم نختار منها بطاقة بطريقة عشوائية حتى نسحب حجم العينة المطلوب.

علماً بأن احتمال اختيار أية بطاقة يساوي $\frac{1}{10}$

ومن الجدير بالذكر أنه يصعب استخدام الطريقة المذكورة سابقاً إذا كان حجم المجتمع كبيراً وفي هذه الحالة يستخدم الباحث جدول الارقام العشوائية (Table of Random Numbers). كما في المثال التالي:

الطريقة الثانية: عندما يكون حجم المجتمع كبيراً

مثال: إذا أردنا الاختيار بطريقة المعاينة العشوائية البسيطة عينة مكونة من 20 طالباً من مجتمع الجامعة المكون من 800 طالباً. نتبع الخطوات التالية:

- أعط كل عنصر من عناصر مجتمع الدراسة رقماً متسلسلاً من N-1، حيث N هو حجم مجتمع الدراسة، نعطي أرقاماً من 001، 002، 000، 050، 600، 800 علماً بأن الرقم المتسلسل لكل فرد يتكون من عدد المنازل لحجم المجتمع وفي هذا المثال لثلثه منازل.
- نختار عشرين رقماً من جدول الأرقام العشوائية ونبدأ من العمود الأول في اليسار ونقرأ ونتجه عمودياً إلى الأسفل بحيث ننظر إلى ثلاثة منازل من جهة اليسار، فإذا كان العدد الذي نقرؤه ضمن أرقام المجتمع لأننا نأخذ ذلك الرقم في العينة وغير ذلك نرفضه.

جدول الأرقام العشوائية Table of Random Numbers

0045	2222
0123	3101
1456	8954
9847	4321
3210	0239
8140	9999
0010	8855
1112	8133
7000	8010
9988	2357

الجدول أعلاه يمثل جزء من جدول الأرقام العشوائية، كما ذكرنا سابقاً نبدأ بقراءة الجدول فنقرأ ثلاث منازل من كل عدد ونأخذ فقط الأعداد الواقعة ضمن أرقام المجتمع ولا نكرر أي عدد أخذناه سابقاً، فتكون العينة هي الأرقام التي تحمل 004، 012، 145، 222، 310، 432، 023، 235 والأرقام التالية:

نرفضها (لا ناخذها) 984، 814، 998، 895، 999، 885، 813، 801، ونستمر في قراءة الجداول حتى يكون حجم العينة العدد المطلوب وهو (20). ونلاحظ بأن جميع الارقام المرفوضة في الجدول هي اكبر من حجم المجتمع وحسب المثال هو 800.

2) العينة العشوائية الطبقية: Stratified Random Sample:

يتم تقسيم المجتمع الإحصائي (غير المتجانس) إلى مجموعات جزئية متجانسة في بعض الخصائص وتسمى كل مجموعة طبقة ومن ثم يتم أختيار عينة عشوائية من كل طبقة لتشكل معاً عينة طبقية.

خطوات اختيار العينة العشوائية الطبقية:

- تحديد الطبقات بوضوح ووضع كل وحدة معاينة من المجتمع في الطبقة الملائمة.
- بعد تقسيم وحدات المعاينة في طبقات وتحديد حجم كل طبقة بعد ذلك تحدد حجم العينة في كل طبقة ومن ثم نختار عينة عشوائية بسيطة من كل طبقة ويتم ذلك بإحدى الطرق التالية:

أ-التخصص المتساوي:

حيث يتم اختيار عدد متساو من الطبقات المختلفة وعلى فرض أن عدد الطبقات (K=4).

وأحجامها (($N_1 = 1000$، (N_2=2000)، = (N_3=3000)، 4000 - N_4))،

وحجم العينة المطلوب (n=200)، فإن حجم العينة = $n_1 = n_2 = n_3 = n_4 \frac{200}{4}$ =50

وهذا يعني أن حجم العينة المطلوب من طبقة يساوي 50.

ب-طريقة النسب:

$n_r = n_t = \frac{N_1}{N} \times n$ حيث أن (n) حجـم العينـة، (N_1) حجـم الطبقـة، (N) حجـم جميع الطبقات. وفي المثال السابق فإن حجم العينة في كل طبقة يكون:

حجم الطبقة الأولى= N_1

حجم الطبقة الأولى = n1

حجم لعينة المطلوبة =n

$$n_1 = \frac{1000}{10000} \times 200 = 20$$

$$n_2 = \frac{1000}{10000} \times 200 = 40$$

$$n_3 = \frac{1000}{10000} \times 200 = 60$$

$$n_4 = \frac{1000}{10000} \times 200 = 80$$

ويكون حجم العينة في كل طبقة هو 20 من الطبقة الاولى، 40 مـن الطبقـة الثانيـة، 60 من الطبقة الثالثة، 80 من الطبقة الرابعـة، ويكـون مجموعـة العينـة الكـلي يسـاوي حجم العينة المطلوب وهو 200.

مثال آخر:

إذا اردنا اختيار عينة حجمها 20 طالباً من مجتمع الجامعة المكون من أربع طبقـات وعلى النحو التالي:

الطبقة الاولى الطبقة الثانية الطبقة الثالثة الطبقة الرابعة

طلاب السنة الاول طلاب السنة الثانية طلاب السنة الثالثة طلاب السنة الرابعة 40

100 200 300

الحل:

حجم العينة المطلوب (n=20)

حجم المجتمع الكلي N يساوي:

$$N = 400 + 300+200+100 = 1000$$

حجم العينة من طبقة السنة الاولى: $n_1 = \frac{1000}{10000} \times 20 = 8$

حجم العينة من طبقة السنة الثانية: $n_2 = \frac{1000}{10000} \times 20 = 6$

حجم العينة من طبقة السنة الثالثة: $n_3 = \frac{1000}{10000} \times 20 = 4$

حجم العينة من طبقة السنة الرابعة: $n_4 = \frac{1000}{10000} \times 20 = 2$

ويكون حجم العينة من جميع الطبقات يساوي:

$n = 8+6+4+2 = 20$ وهو العدد المطلوب.

3) العينة المنتظمة Systematic Sampling

اختيار عنصر بطريقة عشوائية من أول (k) من العناصر في إطار المعاينة ومن ثم اختيار كل عنصر رقمه (k) بعد العنصر المختار سابقاً.

مثال:

إذا اردنا أخذ عينة حجمها 15 من زبائن محل تجاري للألبسة علماً بأنه من الصعب تحديد حجم المجتمع الإحصائي (زبائن المحل) وفي هذه الحالة تستخدم المعاينة المنتظمة ونتبع الخطوات التالية:

1-نختار رقماً عشوائياً من 1-10 مثلاً وليكن العدد 7 فيعتبر هذا العنصر الأول في العينة.

2-نقابل الشخص السابع عندما يخرج من المحل.

3-نقابل الأشخاص الذين تكون أرقام خروجهم 14، 21، 28، 35... وهكذا حتى نحصل على العينة التي حجمها 15.

إن هذه الطريقة تعتبر سهلة بالنسبة إلى شخص غير متدرب، وهي أيضاً موزعة بشكل أكثر تجانساً على جميع أفراد المجتمع.

4) العينة العشوائية المنتظمة Systematic Random Sample:

يتم فيها اختيار الوحدة الأولى بطريقة عشوائية واختيار هذه الوحدة يحدد اختيار بقية الوحدات للعينة حسب فترة معاينة أو فترة الانتظام أو طول الفترة.

مثال:

إذا أردنا اختيار عينة حجمها (n=100) طالباً من مجموع طلاب تخصص إدارة الأعمال وعددهم (N=400) طالباً وذلك لندرس معدلاتهم في مادة مبادئ الاحصاء. بطريقة المعاينة المنتظمة واستخدام طول القفزة:

الحل:

نستخرج طول القفزة: $\frac{N}{n} = \frac{400}{100} = 4$

نسحب رقماً عشوائياً من 1-4 وليكن 3 ومن ثم نضيف على رقم 3 طول القفزة 4 ونسحب البطاقة ذات رقم 7 ثم البطاقة 11، 15، 19، 23 وهكذا حتى آخر بطاقة نسحبها ويكون رقمها 400.

مع ملاحظة التالي : إذا لم يكن طول القفزة عدداً صحيحاً فإننا نقرب الجواب إلى أقرب عدد صحيح.

5) العينة العشوائية العنقودية: Cluster Random Sample

تعتبر المعانة العشوائية العنقودية معاينة ذات مرحلتين والهدف منها تقليل النفقات المادية مع الاحتفاظ بخصائص ومميزات المعاينة الاحتمالية. وحيث يتم تقسيم المجتمع إلى مجموعات جزئية ولا يشترط تجانسها وهذه المجموعات الجزئية تقسم إلى مجموعات جزئية أخرى... ومن ثم نختار عينة عشوائية بسيطة من كل مجموعة جزئية لتشكل في النهاية عينة عنقودية. مثل تعداد السكان: فإذا أراد الباحث اختيار عينة من مدينة الزرقاء مثلاً للتعرف على الأوضاع الاقتصادية للأسر فيها فإنه لا يستطيع استخدام طريقة المسح الشال بسبب ارتفاع التكاليف ففي هذه الحالة يلجأ إلى العينة العشوائية النقودية ويقوم بتقسيم المدينة إلى مناطق ومن ثم أحياء ومن ثم شوارع وبعد ذلك يقرر حجم العينة من منطقة أو حي أو شارع.

ثانياً: العينات غير الإحتمالية: No Probability Sampling

يعتمد اختيار العينة غير الاحتمالية على التقدير الشخصي للباحث ومن نواع العينات:

1) المعاينة بالاختيار السهل (العينة المريحة): Covenience Sample

يختار الباحث العينة التي تحتاج إلى أقل كلفة ووقت وجهد ومن السهل الوصول إليه والحصول على المشاهدات. فمثلاً أراد باحث القيام بدراسة عن محلات تجارية فيقوم الباحث باختيار مفردات العينة أو الجزء الأكبر منها في المدينة أو المنطقة القريبة من سكنه.

مثال آخر:

إذا أراد مراقب الجودة في مصنع ما أن يعطي تقريراً عـن جـودة الإنتـاج في المصـنع وكن يأخذ وحدة من كل صندوق وتفحصها وهكذا يعمل لباقي الصناديق فإن العينـة لا تمثل المجتمع تمثيلا صحيحاً.

2) العينة الفرضة: Purposive Sampling

يختار الباحث العينة بناءً على حكم الشخصي حيث تتصف مفردات العينة بخـواص معينة وتحقق أغراضاً محدودة. فمثلاً يختار صاحب محل تجـاري عينـة مـن الأسر التـي تتعامل مع محله باستمرار بحيث يستطيع تحليل حجـم المشـتريات لهـذه الأسر وانـواع، ومن الممكن أن تعطي هذه الطريقة نتائج إيجابية وذلك إذا كـان رأي الـدارس وحكمـه مناسبين.

3) عينة الحصص: Quota Sample

تستخدم للتأكد من أن جميـع طبقـات المجتمـع وخواصـه ممثلـه في العينـة، حيـث يخصص لجامع البيانات في معاينـة الحصـص عـدد مـن المشـاهدات عليـه أن يقابلهـا أو يجمع البيانات عنها. مثلاً قد يخصص لجامع البيانات في دراسة لقياس نتائج الامتحانـات 50 طالباً منهم 25 طالباً لكلية الاقتصاد 10 لكلية الهندسة، 15 لكلية الحقوق.

تصنيف وتبويب البيانات: Classification and Tabulation of Data

بعد جمع البيانات سواء كانت من المصادر المباشرة وغير المباشرة ومراجعتهـا يحتـاج الباحث إلى تصنيف وتبويب هذه البيانات بشكل يمكن فهمه أو إجراء الحسـابات عليـه، أما التصنيف فهو عبارة عـن تجميـع الحقـائق أو الخصـائص المشـتركة في مجموعـات أو تصنيفات أو فئات مثل ذكـور وأنـاث أو طـلاب سـنة أولى وطـلاب سـنة ثانيـة أو يمتلـك رخصة قيادة ولا يمتلك رخصة قيادة.. أما التبويب فهو إجراء لتلخيص البيانـات وتقـديمها بجداول إحصائية من أجـل تنظـيم البيانـات في أعمـدة وصـفوف حيـث الأعمـدة ترتـب عمودياً والصفوف ترتب أفقياً ويهدف التصنيف إلى:

- تلخيص البيانات وإظهار الخواص المشتركة أو الاختلافات في عد محدد من الفئات.

- تسهيل عملية المقارنة بين الظواهر أو المتغيرات المختلفة.
- تهيئة البيانات لعمليات الحسابات والتحليل الإحصائي.
- إظهار الصفات الهامة وحذف الأمور غير الهامة ويمكن تصنيف البيانات حسب التالي:

1-التصنيف الجغرافي – حسب المواقع الجغرافية مثل تصنيف سكان الأردن حسب المحافظات.

2-التصنيف الزمني – حسب الفترة الزمنية مثل معرفة سعر صرف الدولار خلال السنوات 2000 حتى 2005.

3-التصنيف النوعي (الوصفي) – حسب السمات أو الصفات مل تصنيف السكان حسب الجنس (ذكر/أنثى)، أو (متعلم/ غير متعلم).

4-التصنيف الكمي – حسب كمية كل مجموعة مثل تصنيف 50 أسرة حسب عدد أفرادها.

التوزيعات التكرارية: Frequency Distribution

(تلخيص البيانات وعرضها بطريقة التوزيعات التكرارية)

ولها صورتين:

الاولى: عندما يكون المدى أصغر.

الثانية: عندما يكون المدى كبير.

عندما نقوم بدراسة أو البحث بمشكلة أو ظاهرة ما نعمل على جمع أعداد كبيرة من البيانات ولا يمكن المقارنة بين مفردات هذه البيانات وفهمها لذلك نلجأ إلى تلخيص وعرضها بطريقة مبسطة تسهل علينا فهمها ومن هذه الطرق التوزيعات التكرارية التي تمكنا من تنظيم البيانات الكثيرة بحيث لا تخسر هذه البيانات من أهميتها شيئاً.

الطريقة الأساسية لبناء جدول (بناء جدول التوزيعات التكرارية) عندما يكون المدى صغيراً، التوزيع التكراري: عبارة عن تقسيم مدى قيم البيانات إلى فئات وحصر عدد البيانات الواقعة ضمن كل فئة.

مثال:

الارقام التالية تمثل علامات 20 طالباً في امتحان مبادئ الاحصاء والعلامة القصوى 25 والمطلوب عرض هذه البيانات في توزيع تكراري.

23	18	20	8	25
20	18	15	10	8
18	23	18	15	12
15	12	20	15	18
10	10	15	18	20

الحل:

يمكن تفريغ البيانات في التوزيع التكراري على النحو الآتي:

جدول التوزيع التكراري لعلامات الطلبة في مبادئ الإحصاء

العلامة	التكرار
8	2
10	3
12	2
15	5
18	6
20	4
23	2
25	1
المجموع	25

ونلاحظ أن مدى هذه البيانات يساوي أعلى مشاهدة ناقصاً أدنى مشاهدة ويساوي 17=8-25 وبما أنه لا يوجد طلبة حصلوا على العلامات التالية: 9، 11، 13، 14، 16، 17، 19، 21، 22، 24. فإنه لا يوجد ضرورة إلى الإشارة لها في الجدول حيث تفهم ضمناً من الجداول بأنه لا يوجد طلبة حصلوا عليها.

لذلك نلاحظ في بناء التوزيع التكراري اننا بدانا من أدنى قيمة (علامة) وهي 8 ثم رتبنا القيم تصاعدياً حتى وصلنا إلى أعلى قيمة وهي 25 كما في

العمود الأول. أما عناصر العمود الثاني فتمثل عدد الطلبة الذين تكررت فيها نفس العلامة وأما القيمة التي لم تظهر فإن تكرارها صفراً.

بناء التوزيع التكراري عندما يكون المدى كبيراً:

إذا كان المدى كبيراً أو عدد البيانات كبيراً فلا بد من تقسيم قيم البيانات إلى فئات يتراوح عددها مابين 5، 15، فئة وحسب حجم البيانات وبعد تقسيم قيم البيانات إلى فئات تفرغ البيانات على الفئات وتجمع التكرارات المقابلة لكل فئة. وحتى نقوم بناء جدول تكراري نأخذ البيانات في المثال التالي:

مثال:

البيانات التالية تمثل الأجر الأسبوعي لأربعين عاملاً في مصنع زيوت والمطلوب بناء توزيع تكراري بإستخدام الفئات لهذه البيانات.

35، 22، 33، 47، 50، 45، 42، 21، 15، 17

31، 32، 37، 41، 43، 27، 13، 10، 26، 25

30، 20، 21، 22، 27، 28، 29، 19، 27، 29

39، 33، 38، 37، 27، 25، 38، 31، 40، 39

الحل: لعرض هذه البيانات في توزيع تكراري ذي فئات متساوية نتبع الخطوات التالية:

أولاً: نحدد عدد الفئات ويفضل أن لا تقل عن 5 ولا يزيد عن 15 وهذا يعتمد على حجم البيانات وهنا نفرض أن عددها 7.

ثانياً: نجد مدى البيانات ويساوي أكبر قيمة – أدنى قيمة

$$50-10 = 40$$

ثالثاً: نجد طول الفئة ويساوي $\frac{\text{المدى}}{\text{عدد الفئات}}$

$$\frac{40}{7} = 5.71$$

يقرب طول الفئة إلى عدد صحيح (لا يوجد كسر) ونقرب العدد 5.71 إلى الأعلى فتكون طول الفئة C=6

رابعاً: تحديد معالم الفئة الأولى:

الحد الأدنى للفئة الأولى = أدنى قيمة من البيانات = 10

الحد الأعلى للفئة الأولى = الحد الأدنى + طول الفئة – 1

10+ 6 – 1 = 15

وبناءً عليه تكون معالم الفئة الأولى هي: 10 – 15

ونحصل على بقية الفئات بزيادة طول الفئة للحد الأدنى:

جدول التوزيع التكراري لأجور 40 عاملاً

التكرار fi	الاشارات	حدود الفئات
3	////	10-15
5	~~////~~	16-21
9	//// ~~////~~	22-27
9	~~////~~ //	28-33
7	~~////~~ //	34-39
5	~~////~~	40-45
2	//	46-51
40		المجموع

خامساً: ويمكن تعين الحدود الفعلية للفئة الاولى من خلال طرح نصف وحدة دقة (أي نصف وحدة من الوحدات التي قربت إلى الأعداد في البيانات وعليه فالحد الأدنى الفعلي للفئة الأولى يكون 9.5 = 10 – 0.5 كما نعين الحد الأعلى الفعلي للفئة الأولى وذلك بإضافة طول الفئة للحد الادنى فيكون الحد الفعلي الأعلى للفئة الأولى هو. 9.5 + 6 = 15.5

ونستطيع فعل الشئ نفسه لباقي الفئات.

سادساً: نعين مراكز الفئات Xi نجد مركز الفئة بقسمة مجموع حدي الفئة على 2 ولذلك فمركز الفئة الأولى هو $\frac{10+15}{2} = 12.5$ ويمكن إيجاد مركز أي من الفئات الأخرى بإضافة طول الفئة إلى مركز الفئة التي قبلها فيكون مركز الفئة الثانية هو 18.5 = 6 + 12.5 وهكذا.

ويكون مجموع التكرارات يساوي عدد البيانات n ويجب أن يكون:

$$\sum fi = n$$

وسوف ننشأ جدول جديد يحتوي على الحدود الفعلية للفئات ومركز الفئة.

الحدود الفعلية للفئة	مركز الفئة	التكرار *fi*	حدود الفئة
9.5-15.5	12.5	3	10-15
15.5-21.5	18.5	5	16-21
21.5-27.5	24.5	9	22-27
27.5-33.5	30.5	9	28-33
33.5-39.5	36.5	7	34-39
39.5 - 45.5	42.5	5	40-45
45.5-51.5	48.5	2	46-51

$$\sum \mathbf{fi} = n$$

مع ملاحظة التالي عند عرض التوزيع التكراري لا نكتب عمود إفراغ البيانات (الإشارات).

2-10 التوزيع التكراري النسبي Relative Frequency Distribution:

إن التكرار النسبي لكل فئة هو نسبة تكرار تلك الفئة إلى مجموع التكرارات $p = \frac{f}{n}$

حيث (n) مجموع التكرارات، (f) تكرار الفئة.

اما بالنسبة للتكرار المئوي للفئة فهو التكرار النسبي مضروب 100%. وبـالرجوع إلى البيانات في الجدول السابق نكون جدول جديد كمايلي:

الجدول: التوزيع التكراري النسبي والمئوي لأجور 40 عاملاً.

التكرار المئوي	التكرار النسبي	التكرار	حدود البيانات
3/40×100%=7.5%	3/40=0.075	*3*	10-15
5/40×100%=12.5%	5/40=0.125	*5*	16-21
9/40×1005=22.5%	9/40=0.225	*9*	22-27
9/40×100%=22.5%	9/40=0.225	*9*	28-33
7/40×100%=17.5%	7/40=0.175	*7*	34-39
5/40×100%=12.5%	5/40=.125	*5*	40-45
2/40×100%=5%	2/40=0.05	*2*	46-51
=100%	$\frac{40}{40} = 1$	40	المجموع

- مجموع التكرارات النسبية يجب ان يساوي 1 صحيح.
- ومجموع التكرارات المئوية يجب أن يساوي 100%.

التوزيع التكراري المتجمع:

نحتاج في كثير من الاوقات معرفة عدد المشاهدات التي تزيد أو تقل عن قيمة معينة وتكون التوزيعات التكرارية المتجمعة على شكلين وهما:

أ-التوزيع التكراري المتجمع الصاعد.

ب-التوزيع التكراري المتجمع الهابط.

أ-التوزيع التكراري المتجمع الصاعد: ك

عبارة عن تجميع تكرارات الجدول الاصلي بدءاً بتكرار الفئة الأولى وانتهاء بتكرار الفئة الأخيرة وبالاستعانة بالمعلومات من الجدول السابق نكون جدول التوزيع التكراري المتجمع الصاعد والهابط.

جدول توزيع تكراري متجمع صاعد لأجور 40 عاملاً.

فئات متجمعة صاعدة	تكرارات متجمعة صاعدة	تكرارات نسبية متجمعة صاعدة
أقل من 10	0	0
أقل من 15	3	0.075
أقل من 21	8	0.20
أقل من 27	17	0.425
أقل من 33	26	0.650
أقل من 39	33	0.825
أقل من 45	38	0.95
أقل من 51	40	1.0

جدول توزيع تكراري متجمع هابط لأجور 40 عاملاً.

فئات	تكرارات متجمعة هابطة	تكرارات نسبية متجمعة هابطة
أقل من 10	40	1.0
أقل من 15	37	0.925
أقل من 21	32	0.8
أقل من 27	23	0.575
أقل من 33	14	0.35
أقل من 39	7	0.175
أقل من 45	2	0.05
أقل من 51	0	0

فإذا اردنا أن نعرف عدد العمال الذين يحصلون على أجـر اسـبوعي قـدره 33 دولار وأقل، ننظر إلى جدول المتجمع الصاعد ويكون الجواب 26 عاملاً.

أما إذا أردنا أن نعرف عدد العمال الذين يحصلون على أجـر أسـبوعي أكـبر مـن 33 دولار ننظر إلى المتجمع الهابط ويكون الجواب 14 عاملاً.

الجداول المقفلة والمفتوحة:

إن الجداول التكرارية التي تكون فئاتها قيماً مفـردة أو فـترات مغلقـة مـن الطـرفين تسمى جداول مغلقة وتستعمل في عـرض البيانـات مختلـف أنواعهـا ومـن ميزاتهـا أنهـا محددة ونستطيع إيجاد مراكزها وبالتالي نستطيع إجراء الحسابات الإحصائية عليها مثـل مقاييس النزعة المركزية ومقايس التشتت وغيرها. أما الجداول المفتوحة فهي تلـك التـي يكون فيها فئة أو أكثر مفتوحة معنى غير مغلقة من أحد طرفيهـا أو كلـيهما. وغالبـاً مـا تكون الفترة المفتوحة أما الفئة الأولى أو الفئة الأخيرة مثل جداول الدخل أو العمر..

جدول مغلق من الطرفي

الفئات	التكرار
10-14	3
15-19	4
20-24	2
25-29	1

جدول مفتوح من الطرفين

الفئات	التكرار
أقل من 10	3
15-19	4
20-24	2
أكثر من 25	1

جدول مفتوح من الأسفل

الفئات	التكرار
أقل من 10	3
15-19	4
20-24	2
25-29	1

جدول مفتوح من الأعلى

الفئات	التكرار
10-14	3
15-19	4
20-24	2
أكثر من 25	1

الجداول المنتظمة وغير المنتظمة:

الجداول المنتظمة:

هي تلك الجداول التي تكون أطوال الفئات فيها متساوية مثل الجداول التي شرحنا عنها سابقاً.

الجداول غير المنتظمة:

هي تلك الجداول التي تكون فيها أطوال الفئات غير متساوية وذلك أن التكرارات صغيرة ويوجد تباعد في البيانات. مثلاً إذا كانت أعلى قيمة 1500 وأدنى قيمة 80 والمطلوب تكوين جدول بـ 6 فئات:

طول الفئة يكون $\frac{1500-80}{6} = 237$

في هذه الحالة من غير المنطق تكوين جدول بهذا الطول للفئات .

لذلك نستخدم جدولاً غير منتظم وخاصة للفئة الاخيرة ويكون الجداول كمايلي:

الفئات	التكرار
80-280	3
281-482	1
482-682	5
683-883	2
884-1500	1

تمارين الفصل الثاني

1- عرف علم الإحصاء ووضح المراحل التي يمر بها.
2- وضح كل من الإحصاء الوصفي والإحصاء الاستدلالي.
3- أذكر خطوات البحث العلمي ووضح إحداها.
4- أذكر الفرق بين مصادر جمع البيانات المباشرة وغير المباشرة.
5- من طرق جمع البيانات طريقة المسح الشامل وطريقة العينة، وضخ الفرق بينهما.
6- إذا أراد باحث أخذ عينة حجمها 150 طالباً من كليات الهندسية، الاقتصاد، العلوم والتي عدد الطلبة فيها على التوالي 1200، 2000، 800 فما عدد أفراد العينة من كل كلية.
7- ماطول القفزة لعينة عشوائية منتظمة حجمها 100، علماً بأن حجم المجتمع الإحصائي 1000.
8- البيانات التالي تمثل أطوال 32 طالباً بالسنتمتر والمطلوب توزيع هذه البيانات في جدول تكراري ذي فئات متساوية. والبيانات كمايلي:

175، 150، 149، 180، 150، 200، 165، 160
162، 140، 135، 190، 162، 148، 175، 179
178، 200، 210، 152، 170، 185، 181، 163
220، 215، 138، 141، 167، 169، 181، 174

9-حسب معطيات السؤال السابق، أوجد التوزيع التكراري النسبي والتوزيع التكراري المتجمع الصاعد والهابط.

10-الجدول التالي يمثل علامات 40 طالباً والعلاقة القصوى 29.

فئة العلاقة	عدد الطلاب
5-0	5
11-6	10
17-12	15
23-18	8
29-24	2

المطلوب:
أ-تمثيل هذه البيانات بيانياً بأسلوب المدرج التكراري.
ب-تمثيل هذه البيانات بالقطاع الدائري.

الفصل الثالث

جمع وعرض البيانات

تصميم البحث العلمي يتطلب خطة تحتوي على العناصر الأساسية المتعلقة بمشكلة البحث، أول هذه العناصر هي تصميم مشكلة البحث وصياغتها بصورة واضحة مع تحديد المنهج الذي يجب اتباعه، وينقسم منهج البحث العلمي عادة إلى قسمين:

1. المنهج النموذجي وهو يحدد الإطار النظري للبحث في استنباط الأهداف واختبار الفروضات.
2. المنهج التطبيقي فهو عبارة عن تطبيق المنهج النموذج باستخدام المفاهيم الإحصائية لجمع البيانات وتحليلها.

وهذا يتطلب تحديد نوع البيانات ومصادرها وما إذا كانت بيانات شاملة أو بيانات معاينة وكيفية اختيار العينات. ويتطلب أيضا وضع خطة للعرض الجدولي والبياني وطرق التحليل واستخلاص النتائج وتعتبر عملية جمع البيانات من أهم العمليات الإحصائية وأكثرها حساسية وتحتاج إلى دقة كبيرة، وتتوقف عليها جميع النتائج الخاصة بالبحث، فكلما كانت هذه العملية دقيقة كلما كانت النتائج التي يتم التوصل إليها دقيقة أيضا والعكس صحيح.

مصادر البيانات Data Sources

البيانات الإحصائية لها مصدران أساسيان:

1- المصادر التاريخية (أو الثانوية)

وهي مصادر غير مباشرة وتشمل الإحصاءات المنشورة والكتب والمطبوعات والوثائق التي تصدرها الحكومات والهيئات الرسمية وشبه الرسمية، وكذلك المنظمات الدولية كالأمم المتحدة ومنظمة العمل الدولية وغيرها، وتقوم هذه الجهات بحكم اختصاصها بجمع البيانات وجدولتها وإصدارها في نشرات دورية من الأمثلة على ذلك الكتاب الديمغرافي السنوي الذي تصدره الأمم المتحدة عن التعدادات العامة للسكان في مختلف دول العالم والتقارير التي تصدر عن الجهاز المركزي للإحصاء حول السكان في السودان.

والمصادر التاريخية (الثانوية) يعاب عليها أنها ربما تحتوي على أخطاء مقصودة أو غير مقصودة كالأخطاء الناتجة عن عدم فهم صاحب المصدر الثانوي لما أخذه من المصدر الأصلي أو أخطاء بسبب تحيز صاحب البحث الثانوي مثل ذكر جزء وترك جزءاً آخر أو إغفال التعاريف أو الشروط الواردة في المصدر الأصلي، أو أخطاء بسبب تحيز صاحب البحث الثانوي مثل ذكر جزء وترك جزء آخر أو و إغفال التعاريف أو الشروط الواردة في المصدر الأصلي، ولذلك فإن المصدر الأصلي أفضل كثيرا من المصدر الثانوي.

2- المصادر الميدانية (أو الأولية)

وهي مصادر مباشرة وتشمل الوحدات الأصلية التي تستقصي منها المعلومات بصفة مباشرة، ويتم جمع البيانات طبقا لهذه المصادر عن طريق:

1) المشاهدة: ففي حالة بحث عدد المدارس والمستشفيات مثلا يمكننا جمع بيانات بمجرد المشاهدة دون مقابلة الأفراد، ويتم جمع البيانات طبقا لهذه الوسيلة في استمارة بحث عن طريق المشاهدة.

2) المقابلة: ويتم جمع البيانات طبقا لهذه الوسيلة في استمارة استبيان عن طريق المقابلة حيث يقوم الباحث بمقابلة الأشخاص ويوجه الأسئلة إليهم ثم يدون الإجابة، ومن مميزات المصادر الميدانية أن الباحث يستطيع التحقق من صحة المعلومة وأن كان يعاب عليها كثرة ما تتطلبه من نفقات.

3) المراسلة: وطبقا لهذه الوسيلة يعد الباحث استمارة استبيان يرسلها بالبريد إلى الأشخاص المبحوثين على عناوينهم للإجابة على الأسئلة التي تضمنتها الاستمارة ثم إرسالها إلى الباحث بعد ذلك، وعلى الرغم من أن هذه الطريقة تؤدي إلى انخفاض التكلفة إلا أنها ليست شائعة الاستخدام نسبة عدم التجاوب العالية كما أنها غير مجدية بالنسبة للمجتمعات التي لا تعرف القراءة والكتابة.

4) الهاتف: وفيها يتم الاتصال بالوحدة عن طريق الهاتف وأخذ المعلومات منها.

وقد أتاحت التقنيات الحديثة طرقاً أخرى لجمع البيانات منها إستخدام الانترنت والفاكس.

أساليب جمع البيانات: Methods of Data Collection

هنالك طريقتان لجمع البيانات الإحصائية وهما طريقة الحصر الشامل وطريقة المعاينة.

الحصر الشامل: Full Count

ويعني أخذ المعلومات من جميع مفردات المجتمع المراد دراسته. ومن أهم خصائصه الشمول الكامل بالمستوي الزمني والمكاني لجميع مفردات الإطار. والإطار يقصد به تمثيل بشكل ما لجميع مفردات المجتمع التي تخضع للدراسة، فمثلا إذا كان المجتمع المراد دراسته هو طلاب كلية الطب بجامعة الجزيرة فإن الإطار يمكن أن يكون – مثلاً – قائمة بأسماء هؤلاء الطلاب.

أسلوب الحصر الشامل يمكن استخدامه إذا كانت مفردات المجتمع محدودة أو في الحالات التي تتطلب جمع البيانات عن خصائص المجتمع على المستوى القومي، مثل ذلك التعداد العام للسكان والمنشآت والتي تقوم به مصلحة الإحصاء على فترات زمنية منتظمة.

من مزايا الحصر الشامل الإلمام بجميع البيانات الخاصة بالمجتمع ككل كما يمكن من إجراء جميع الدراسات المطلوبة كما يؤدي إلى دقة النتائج وعدم تحيزها، ويعاب على أسلوب الحصر الشامل أنه يكون عرضة أخطاء الشمول الناتجة عن تكرار أو نقص الحصر إلى جانب أخطاء المحتوي والتي تصيب بعض خصائص المفردات بالإضافة إلي كثرة النفقات وطول الوقت المطلوب لإجراء البحث واستخراج النتائج.

المعاينة: Sampling

وفي هذه الطريقة يتم جمع البيانات باستخدام عدد محدود من مفردات الإطار ثم نعمم نتائج الدراسة على جميع مفردات المجتمع وتعتبر المعاينة الإحصائية من الوسائل الأكثر استخداما لأنه وباتباع الطرق والنظريات الإحصائية العلمية يمكن معها تعميم النتائج فيما بعد على المجتمع.

ومن مزايا طريقة المعاينة السهولة والبساطة واختصار الوقت والتكاليف كما يؤدي التحكم في عدد محدود من مفردات الإطار إلى نتائج ربما تكون أكثر دقة من الحصر الشامل، إلا أن من عيوب طريقة المعاينة هي أنه إذا

كانت العينة غير ممثلة للمجتمع فإن هذا يؤدي إلى أخطاء التحيز بحيث تصبح نتائجها غير مطابقة تمام كما لو تمت الدراسة باستخدام التعداد الكامل كما سنرى فيما بعد.

وربما يتوجب على الباحث أن يستخدم طريقة المعاينة في الحالات التي يؤدي فيها الشمول الكامل إلى فقدان المجتمع بأكمله، مثل لذلك عندما يفحص الطبيب الدم لمريض فإنه يكون مضطرا لأخذ عينة من الدم لأنه إذا أخذ الدم كله لمات ذلك المريض، أو إذا أراد أحد الباحثين الزراعيين دراسة نوع معين من أمراض الجذور لنوع من أنواع النباتات فلابد له أن يأخذ عينة وإلا فقد النبات بأكمله.

أولاً: المفهوم الإحصائي للمعاينة:

كما ذكرنا سابقا فإن نظرية العينات تقوم على أساس دراسة خصائص مجتمع معين عن طريق دراسة عينة، تمثل ذلك المجتمع.

وفكرة المعاينة هي فكرة مأخوذة من السلوك البشري العام حيث أن فكرة دراسة المجتمع عن طريق دراسة جزء منه هي فكرة شائعة وتمارس يوميا في مختلف مناحي الحياة، فالتاجر يقوم بفحص جزء من السلع المراد شراؤها وصاحب المصنع يقوم بضبط الجودة لإنتاجه بأخذ عينة من الإنتاج وفحصها، لكن مثل هذا السلوك غالبا ما يتم خارج الإطار الإحصائي بافتراض أن مفردات المجتمع تحت الاختبار متجانسة.

على أن المفهوم الإحصائي يحدد تحديا قاطعا لكي تحكم على الكل يفحص الجزء لابد من اتباع أسس منظمة وسليمة يتم بها اختبار هذا الجزء، وهذه الطريقة المنظمة تعرف باسم طريقة المعاينة والجزء المختار يسمى عينة، وأهم الأسس التي تقوم عليها نظرية المعاينة هي فكرة الصيغة العشوائية والتي تمثل أحد أساليب جمع البيانات والتي تمثل نظرية متكاملة من النظريات الإحصائية التي تحكمها قوانينها الرياضية.

ثانياً تصميم العينة Sample Design

إذا تعين إجراء بحث عن طريق العينة فإنه يجب على الباحث أن يأخذ في الاعتبار المرتكزات الآتية:

(أ) تعريف المشكلة المراد بحثها، ويعتبر هذا التعريف ضروريا للتعرف على نوع وكمية البيانات المطلوبة.

(ب) تعريف وتحديد المجتمع المراد بحثه، ويبدأ ذلك بتحديد الإطار الذي تسحب منه مفردات العينة.

(ج) تحديد وحدة المعاينة المطلوبة فإذا كان البحث يتعلق بموضوع اقتصادي مثلا فإن أنسب وحدة للمعاينة هي المنشأة الاقتصادية أو الحقل.

(د) تحديد البيانات المطلوبة من أفراد العينة مع تحديد طريقة جمعها سواء كان ذلك عن طريق الاتصال المباشر أو وغير المباشر.

(هـ) تنظيم ومراقبة العمل الميداني وذلك بوضع جدول زمني لمراحل العمل وتوفير العدد اللازم من الباحثين والمشرفين الميدانيين.

(و) ربما احتاج الباحث في بعض الأحيان أن يقوم بمسح تمهيدي للاستفادة منه في تصميم وتحديد حجم العينة والتأكد من سلامة استمارة الاستبيان الموضوعة.

ثالثاً: الإطار: Sample Frame

كما ذكرنا سابقا فإن الإطار يمثل مفردات المجتمع الخاضعة فعلا للمعاينة أي جميع الوحدات الأولية التي يتكون منها المجتمع والتي هي في متناول اليد ولكي يكون الإطار مناسبا يجب أن تتوافر فيه الصفات التالية:

- أن يشتمل بقدر الإمكان على مفردات المجتمع المراد دراسته.
- أن تكون مفرداته مرتبة رقميا وسهلة التمييز وفي نفس الوقت تعكس خصائص المجتمع.

إن أهم شرط من شروط الإطار هو شموله لجميع مفردات المجتمع حتى تتحقق قاعدة إعطاء فرص متساوية لجميع المفردات للظهور في العينة.

رابعاً: وحدات المعاينة: Sample Units

بداية لابد من معرفة الفرق بين وحدة الدراسة ووحدة المعاينة. وحدة المعاينة هي الوحدة التي تبنى على أساسها المعاينة (ماذا نسحب ؟)، أما وحدة الدراسة فهي الوحدة التي تؤخذ منها المعلومات. فإذا أراد أحد الباحثين اختيار عدد من الفصول بكلية التربية لبحث ظاهرة عدم الإكمال في

الامتحانات فإن وحدة الدراسة هي الطالب أما وحدة المعاينة فهـي الفصـل لأن الوحـدة التي ستسحب هي الفصل بينما التي ستؤخذ منها المعلومات هي الطالب.

و وحدات المعاينة قد تكون وحدات طبيعية بسـيطة كـأفراد المجتمـع أو مجموعـة مـن هـذه الوحـدات كـالأسر؛ كـما قـد تكـون وحـدات مصـطنعة كتقسـيم الطـلاب إلى مجموعات واعتبار كل مجموعة وحدة معاينة.

خامساً: طرق المعاينة: Techniques Sampling

هنالك طرق كثيرة للمعاينة يمكن تقسيمها إلى مجموعتين متميزتين:

المعاينة العشوائية Random Sampling

المعاينة العمدية Purposive Sampling

والمعاينة العشوائية هي التي يتم أخذها على أساس عشوائي أو احـتمالي مكـن معـه تعميم نتائج العينة على المجتمع، وكلمة عشوائي هنا تعني إعطاء فرص متكافئة لجميـع مفردات المجتمع عند أخذ العينة بحيث يكون لكل مفـردة احـتمال (موجـب) معـروف ومحدد سلفا.

والعينة العشوائية الممتازة تحتوي على أنـواع المفـردات المختلفـة للمجتمـع بنسـب قريبة جدا من النسب الموجودة لهذه الأنواع في المجتمع الأصلي.

أما المعاينة العمدية فهي التي يـتم أخـذها بطريقـة غـير عشـوائية وتسـمى أحيانـا بالمعاينة الفرضية فيلجأ إليها إذا لم تكن هنالك أي معلومات متوفرة عن المجتمـع المـراد دراسته أو إذا أريد اختيار عينة صغيرة الحجم من مجتمع كبير نسبيا، كذلك نسـتخدم في الحالات التي يـراد فيهـا تكـوين فكـرة سريعـة عـن مشـكلة معينـة أو لتجربـة اسـتمارة الاستبيان في مسح تمهيدي.

- **المعاينة العشوائية: Random Sampling**

المعاينـة العشـوائية تنقسـم إلى عـدة أنـواع حسـب الطريقـة التـي ينفـذ بهـا مبـدأ العشوائية.

(أ) المعاينة العشوائية البسيطة: Sampling Simple Random

إذا أردنـا سـحب عينـة حجمهـا n مـن مجتمـع مـا، فإنهـا توصـف بأنهـا عشـوائية بسـيطة إذا كـان لكـل مجموعـة ذات حجـم n يمكـن سـحبها مـن

ذلك المجتمع نفس الفرصة في أن تكون هي المسحوبة. ويتحقق ذلك عندما يكون لكل مفردة من مفردات المجتمع نفس الفرصة للظهور في العينة، ولتطبيق مبدأ العشوائية نستخدم ما يسمى بالصندوق المثالي أو نلجأ للجداول العشوائية.

مثال (1):

إذا أردنا أن نأخذ عينة عشوائية بسيطة تتكون من 10 طلاب في فصل به 200 طالب فإننا نبدأ بترقيم جميع الطلاب في الفصل أما إذا استخدمنا جداول الأرقام العشوائية فإننا نرقم أفراد المجتمع من 1 إلى 200 كما سبق ثم نستعمل جدول الأرقام العشوائية رقم (2) الملحق بنهاية هذا الكتاب وتقوم بسحب عينة عشوائية بسيطة ذات حجم 10 من هذا المجتمع.

ويمكن عمل ذلك بعدة طرق إحداها أن نقرأ هذا الجدول رأسيا بحيث يكون عدد الخانات أو والأعمدة مساوية لعدد خانات الرقم الذي يمثل مجموع المجتمع محل الدراسة والعدد الذي يكون أصغر من عدد المجتمع نأخذه والذي يكون أكبر من عدد المجتمع نتركه حتى نصل إلى أرقام يكون عددها مساويا لعدد مفردات العينة.

وإذا ما نظرنا إلى الجدول وقرأنا أول ثلاثة أرقام في الصف والعمود الأول نجد 772 يليه رأسيا 033 وهنا نأخذ الثاني ونترك الأول وهكذا نجد أن:

العينة المكونة من 10 طلاب هي:

33	**97**	**136**	**184**	**173**
133	**157**	**154**	**77**	**78**

وهذه الأرقام مأخوذة من العمودين الأول والثاني من الملحق رقم (2) ومن مزايا المعاينة العشوائية البسيطة إنها سهلة الاختيار إلا أن عيبها الرئيسي أنه لا ينصح باستخدامها إلا إذا كان المجتمع الأصلي متجانسا من حيث التركيب والتكوين.

(ب) المعاينة الطبقية:

المعاينة الطبقية يلجأ إليها في حالة معرفة التركيب النسبي للمجتمع الأصلي عندما يكون هذا المجتمع مكونا من عدة طبقات بينها اختلاف

واضح. فقد تعتمد النتيجة الإحصائية في بعض الأحيان على بعض الصفات كالعمر والجنس والمهنة، وفي هذه الحالة يتم تقسيم المجتمع حسب الصفات ثم بعد ذلك يتم تحديد حجم العينة بتخصيص نصيب كل صفة أو طبقة من حجم العينة بحيث نضمن تمثيل العينة لكل طبقات المجتمع.

ويتم توزيع العينة بين الطبقات بثلاث طرق هي:

1. التوزيع المتساوي وفيه يتم توزيع العينة على جميع طبقات الإطار بالتساوي.
2. التوزيع المتناسب وفيه يتم توزيع العينة على طبقات الإطار بصورة تتناسب مع حجم كل طبقة.
3. التوزيع الأمثل وفيه يؤخذ عدم التجانس داخل الطبقات وحجم الطبقة وتكلفة مشاهدة الوحدة فيها في الاعتبار بحيث يزيد عدد المفردات المأخوذة من الطبقات غير المتجانسة وذات الحجم الكبر والأقل تكلفة.

ولكن وفي كل الحالات فإن المعاينة الطبقية يراعى فيها أول ما يراعي أن تكون الطبقات متباينة فيما بينها متجانسة في داخلها وفي هذه الحالة وحدها تكون أخطاء المعاينة للعينة الطبقية أصغر من أخطاء المعاينة العشوائية البسيطة.

مثال (2)

إذا كان الفصل في المثال السابق به 120 طالب و 80 طالبة فإن أخذ عينة طبقية بتوزيع متناسب من هذا الفصل حسب النوع يتم كما يلي:

1. نقسم الفصل إلى قسمين بالتناسب أي 120: 80 أي 120 طالب و 80 طالبة.
2. تؤخذ من كل فئة عينة عشوائية بسيطة بحيث يكون إجمالي حجم العينة يساوي 10 طلبة وطالبات.

ومن ذلك يتضح أننا لابد أن تأخذ 4 من الطالبات و 6 من الطلاب وتتميز المعاينة الطبقية على غيرها لأنها تستهدف دراسة كل طبقة من طبقات المجتمع على حد إلى جانب دراسة المجتمع بأكمله.

وتعطي المعاينة الطبقية أفضل النتائج إذا كانت الصفة التي يتم على أساسها التقسيم الطبقي ترتبط مع الصفة المطلوبة دراستها في العينة.

(ت) المعاينة المنتظمة:

وفيها يتم ترتيب مفردات المجتمع بأرقام مسلسلة ثم تحدد فترة المعاينة بقسمة حجم المجتمع الكلي على حجم العينة. فإذا كان طول هذه الفترة k مثلاً، نختار وحدة عشوائياً من الـ k وحدة الأولى، ثم نختار كل k وحدة بعد ذلك. فإذا كان ترتيب الوحدة المختارة r مثلاً نختار الوحدات ذات الرتب ....,r + 2k , r + k وهكذا.

مثال (3):

تعود إلى المثال السابق ونأخذ عينة منتظمة مقدراها 10 طلاب من عدد 200 طالب. فإذا افترضنا أن حجم المجتمع يساوي (N) وحجم العينة يساوي (n) فإن فترة المعاينة هي:

$$1= N/n = 200/10 = 20$$

فإذا بدأنا بالطالب رقم 1 فإن أرقام الوحدات التي تدخل العينة المنتظمة تكون على النحو التالي :

1	21	41	61	81
101	121	141	161	181

والسؤال هنا هو كيف يتم تحديد رقم المفردة الأولى في المعاينة المنتظمة ؟ فمثلا إذا بدأنا بالطالب رقم 5 فإن المعاينة المنتظمة تكون على النحو التالي:

5	25	45	65	85
105	125	145	165	185

وهكذا فإن اختيار المعاينة المنتظمة يعتمد على اختيار المفردة الأولى وفترة المعاينة، ولتحقيق مبدأ العشوائية في المعاينة المنتظمة يتم السحب اعتباراً من رقم مفردة يتحدد عشوائيا، ويمكن أن يتم ذلك باختيار مفردة واحدة من العشرين مفرده الأولى بطريقة الصندوق المثالي:

ومن مزايا المعاينة المنتظمة السهولة وقلة التكلفة والجهد.

(د) المعاينة متعددة المراحل Multi Stage Sampling

هذا النوع من المعاينة يستخدم في الدراسات الكبيرة ذات الطابع القومي حيث إنه يستفاد من التنظيم الإداري للدولة في تقسيم وحدات المعاينة،

والمعاينة متعددة المراحل يتم فيها عشوائيا اختيار عدد معين من وحدات المعاينة الأولية ثم من داخل كل وحدة يتم اختيار وحدات المعاينة التالية ثم من داخل كل وحدة يتم اختيار وحدات أقل حجما عشوائيا ويمكن تكرار العمل أكثر من مرحلة، على سبيل المثال لقد كان التعداد الأول في السودان لعام 1956 تعداد بالعينة واستخدمت فيه المعاينة متعددة المراحل. كانت المديريات في ذلك الوقت تمثل وحدات المعاينة الأولية وقد تم اختيار عدد من المراكز من داخل كل مديرية بطريقة عشوائية في المرحلة الثانية، في المرحلة الثالثة تم اختيار عدد من المعموديات من داخل كل مركز بطريقة عشوائية. ومن كل عمودية تم اختيار عدد من المشائخ بطريقة عشوائية في المرحلة الرابعة وفي المرحلة الخامسة والأخيرة تمت معاينة جميع الأسر الواقعة في نطاق هذه المشائخ المختارة.

مثل هذا النوع من المعاينة يسمى معاينة من خمس مراحل.

من أهم مزايا المعاينة متعددة المراحل انخفاض تكلفة انتقال الباحثين ففي المثال السابق نجد أن الأسر التي يتم اختيارها بهذه الطريقة تكون متمركزة في أقل عدد من القرى مما يقلل تكاليف دراستها مقارنة بالمعاينة العشوائية البسيطة حيث يتم الاختيار من كل المديرية.

ولكن من عيوب المعاينة متعددة المراحل هو تراكم حجم الخطأ العشوائي ذلك لأن خطأ عدم التجانس يزداد حجما من مرحلة إلى أخرى، لهذا السبب لا بد وأن نتأكد أن المرحلة الأولية تم تقسيمها تقسيما يضمن التمثيل الكافي للمجتمع الأصلي.

(هـ) المعاينة العنقودية: Cluster Sampling

في بعض الأحيان يتم تقسيم المجتمع إلى مجموعات ثم يتم التعامل مع كل مجموعة كوحدة معاينة قائمة بذاتها وتستخدم هذه الطريقة في حالات استبيانات الرأي بصورة منتظمة بحيث أنه وبعد اختيار المرحلة الأولى للمعاينة يمكن إعادة استخدامها بصورة مستمرة لدراسة التغيرات في الرأي العام، وتختلف المعاينة العنقودية عن المعاينة الطبقية في مبدأ العناقيد والذي يحدد أن تكون العناقيد متباينة في داخلها متجانسة فيما بينها.

وفي المعاينة الطبقية يتم الاختيار من داخل الطبقات أما في المعاينة العنقودية يتم الاختيار بين العناقيد، والمعاينة بمرحلتين أو أكثر تعتبر أمثلة للمعاينة العنقودية.

والمعاينة العنقودية تمتاز بالسهولة وانخفاض التكلفة إلا أن عيبها الأساسي يتلخص في زيادة حجم الخطأ الناجم عن تقسيم مجمع غير متجانس.

(و) المعاينة المتعددة الأغراض: Multi Purpose Sampling

في هذه الطريقة يتم سحب عينة كبيرة من مفردات المجتمع لدراسة ظاهرة محددة ثم يتم سحب عينة أصغر من العينة الأولى لدراسة ظاهرة أخرى ثم عينة أصغر من العينة الثانية لدراسة ظاهرة أخرى وهكذا، على سبيل المثال من إطار يشمل جميع النساء نختار عينة لدراسة مستوى الدخل، ثم نختار عينة أصغر من النساء بغرض دراسة ظاهرة العنف ضد المرأة ومن هذا الإطار الجديد يمكن أن نختار عينة أصغر لدراسة النساء المتزوجات فقط بغرض دراسة استعمال وسائل منع الحمل وهكذا. ويلاحظ أن هذه الطريقة للمعاينة تمتاز بثبات وحدة المعاينة في كل المراحل إلا أنها لا تصلح في الحالات التي يتوجب أن يقسم فيها المجتمع إلى طبقات أو مراحل حيث أن وحدات المعاينة تختلف في طبيعتها من مرحلة إلى أخرى.

♦ **المعاينة العمدية: Purposive Sampling**

في شرحنا للمعاينة الاحتمالية وجدنا أن مبدأ الاختيار العشوائي لمفردات العينة يمثل المرتكز الأساسي للعملية الإحصائية.

أما في المعاينة العمدية فإن الباحث لا يهتم كثيرا بمبدأ العشوائية بل وفي بعض الأحيان يتحكم في اختيار أو عدم اختيار المفردة في العينة.

والمعاينة العمدية لا علاقة لها بالعملية الإحصائية إلا من قبيل الصدفة، ذلك لأنه لا يجوز هنا تطبيق قواعد نظرية العينات ويتم إجراء العينة بإحدى طريقتين:

(أ) طريقة الاختيار الوسيط: Average Selection Method

وفيها يختار الباحث مفردات المعاينة من المجتمع تحت الدراسة حسب وجهة نظره على إنها تمثل مفردات وسط، فمثلا في حالة اختيار عينة لذوي

الدخل المحدود ربما يختار الباحث الأشخاص من ذوي الدخل المتوسطة لا هـي بالعاليـة ولا هي بالمنخفضة وبذلك تصبح عينة عميدة بحيث يعتمد عليها في العملية الاختياريـة الوسطية.

(ب)طريقة الزمرة: Quota Sampling

وفي هـذه الطريقـة يـتم الاختيـار بطريقـة صـدفية بحتـة دون الاهـتمام بالأسـلوب العشوائي على سبيل المثال إذا أراد تاجر فواكه أن يفحص صلاحية كمية من البرتقال فإنه ربما يختار صندوق أو صندوقين من هذا البرتقال ويقوم بفحص مفرداته ويقرر على ضوء ذلك سلامة أو عدم سلامة البرتقال.

ونلاحظ في هذه الطريقة أن ذلـك التـاجر قـد قـام باسـتخدام أسـلوب التقسـيم الطبقـي بطريقة عمدية وبأقل تكلفة حيث لا يحتاج الأمر إلى صبر إحصائي لمعرفة أخطاء المعاينة. وقـد ينجح ذلك التاجر في اتخاذ قراره إذا كان على علم مسبق ببعض المعلومات عن ذلك البرتقال.

من كل ما تقدم يتضح لنا جليا أن أسلوب المعاينة هو أمر شائع بين الناس في الحياة العادية، ولكن من المهم أن تفرق بين المعاينة العلميـة القائمـة عـلى الاحـتمال والاختيـار العشوائي والتي تمكن من قياس الأخطاء والتحكم فيها وبين المعاينة غير العشوائية والتـي لا تمكننا من الوصول إلى تقديرات إحصائية بالمعنى العلمي كما لا يترتب عليها إجـراءات لاحقة أو دراسات متأنية.

سادساً أخطاء المعاينة Sampling Errors

كما ذكرنا سابقا أن الغرض من المعاينة هو دراسة الجزء بغرض التعمـيم عـلى الكـل أي تحليل مفردات العينة وتعميم نتائجها على المجتمع، وعند تنفيذ هـذه العمليـة فـإن هنالك نوعان من الأخطاء المحتملة وهي:

أ) الخطأ العشوائي: Random Error

وهو الخطأ الذي ينتج عن طريقة الصدفة الناتجة عـن طبيعـة الاختيـار العشوائي ويتوقف هذا الخطـأ عـلى طريقـة المعاينـة وحجـم العينـة وتبـاين المجتمـع، فكلـما زاد حجـم العينـة أو زادت درجـة التجـانس بـين مفـردات المجتمع يقل مقدار الخطأ العشـوائي، ولـذلك فإنـه يمكـن الـتحكم في الخطـأ

العشوائي بزيادة حجم العينة، كما أن مبدأ العشوائية لو طبق تطبيقا سليما لامكن معه تقدير حجم الخطأ من العينة نفسها.

ب) خطأ التحيز: Bias Error

وهذا الخطأ يحدث عادة بسبب الاختيار غير العشوائي و ينتج عن التحيز عند اختيار المفردات وهنالك أمور كثيرة تؤدي إلى التحيز في الاختيار تجمل أهمها فيما يلي:

1. اختيار مفردات العينة وفق الخاصية المراد قياسها.
2. التحيز الإرادي والذي يكون الهدف منه تحقيق غرض ما في نفس الباحث.
3. عدم وجود إطار سليم لمفردات المجتمع عند إجراء المعاينة.
4. الاستبدال المقصود لبعض وحدات مختارة من العينة لوحدات بديلة غير مختارة.
5. الفشل في الحصول على قسم كبير من المفردات المختارة في العينة.

وخطأ التحيز ليس قاصرا على أسلوب المعاينة فقط وإنما يشمل الحصر الشامل أيضا نتيجة لعدم الدقة في القياس أو عدم الشمول الناتج عن الكم أو المحتوى.

سابعاً: أخطاء غير المعاينة:

وهي أخطاء شائعة تحدث أثناء مراحل جمع وتجهيز البيانات ويمكن أن نجملها فيما يلي:

1. أخطاء التبليغ وتحدث من الباحث بسبب النسيان أو عدم الفهم للمادة المبحوثة (وهذا يكون في التعداد السكاني).
2. عدم التجاوب في التزويد بالمعلومات الكافية مما تؤدي الي خطاء من المبحوث أما لعدم الاهتمام أو لعدم وجود المعلومات الكافية عن المادة المبحوثة.
3. أخطاء الترميز وتحدث عادة في مرحلة تجهيز البيانات عند استبدال الكلمات برموز رقمية غير مطابقة.

عرض البيانات: Data Presentation

العرض الجدولي للبيانات:

بانتهاء عملية جمع البيانات تبدأ عملية المراجعة. والمراجعة قد تكون مكتبية او ميدانية الغرض منها محاولة اكتشاف الأخطاء قبل تصنيف البيانات وتبويبها.

بعد الانتهاء من عملية المراجعة تبدأ عملية تبويب البيانات وذلك بغرض تنظيمها وتلخيصها وإعدادها بصورة يسهل معها استخلاص أهم النتائج المضمنه فيها.

يبدأ عرض البيانات الإحصائية بتصنيفها حيث يلزم التفرقة بين البيانات الوصفية والكمية.

البيانات الوصفية:

هي عبارة عن مجموعة أوصاف لبعض الخصائص حسب ورودها في استمارة الاستبيان. وتنقسم الجداول التي يمكن بواسطتها تبويب البيانات الوصفية إلى نوعين: جداول عامة وأخرى تلخيصية.

يشتمل النوع الأول على كافة المعلومات التي ترد عادة في استمارة الاستبيان وهي جداول تفصيلية في معظم محتوياتها. وتستخدم كمراجع للدراسة, ونجدها غالباً كملاحق في نهاية التقرير للبحث.

ولتوضيح ذلك نعرض المثال التالي

مثال (1):

كانت المتغيرات في نتيجة الإحصاء لـ 50 طالباً في الفصل الدراسي الثالث بقاعدة العلوم الاجتماعية في كلية الاقتصاد والتنمية الريفية للعام الدراسي 1991/90م كما يلي:

D	B	C	A	C	B	A	B	B	A
D	A	D	B	D	A	B	B	C	B
F	F	B	A	C	B	A	C	A	C
F	B	A	B	F	B	B	D	B	B
D	A	C	B	B	A	A	D	B	B

البيانات الواردة في مثال (1) هي بيانات تفصيلية ولا يمكن الاستفادة منها في دراسة, وذلك لعدم وضوحها وصعوبة استنتاج أي معالم عن مستوى الطلاب في الفصل فمثلاً لا يمكننا معرفة عدد الطلاب الذين نالوا تقدير A بسهولة من البيانات أعلاه وخاصة إذا كان عدد الطلاب كبيراً.

لذلك أصبحت الحاجة إلى إيجاد طريقة لتنظيم وتلخيص مثل هذه البيانات في صورة سهلة ضرورية. ولذلك نلجأ إلى ما يعرف بالجداول التلخيصية.

والجداول التلخيصية تشتق عادة من الجداول التفصيلية العامة وتستخدم بغرض إظهار المؤشرات والمقاييس الإحصائية اللازمة لوصف البيانات الواردة في الدراسة واهم أنواع الجداول التلخيصية هو الجدول التكراري.

وسنقوم بتعريف وشرح طريقة عمل الجدول التكراري فيما يلي:

الجداول التكرارية الوصفية:

اولى الخطوات في تكوين الجدول التكراري الوصفي هي تكوين جدول تفريغ البيانات ومنه نستنتج جدولاً اخراً يسمى جدول التوزيع التكراري على النحو التالي؛ انظر جدول رقم (1)

جدول رقم (1)

تفريغ البيانات لتقديرات الطلاب في مثال (1)

التقدير	العلامات	التكرار (عدد الطلاب)
A	//// //// //	12
B	//// //// //// ////	20
C	//// //	7
D	//// //	7
F	////	4
المجموع		**50**

ونلاحظ إننا في العمود الأول نكتب الصفة وهي في هذه الحالة التقدير. أما في العمود الثاني فيكون العنوان العلامات. ونضع لكل قراءات علامة أمام التقديرات الموجودة في العمود الأول. والعلامة عبارة عن خط رأسي مائل (/) فإذا وصل عدد العلامات اى أربعة فان الخط الخامس يكتب بشكل

عرضي لتشكل المجموعة حزمة خماسية. وبعد تفريغ جميع البيانات تعد الحزم أمام كل صفة ويكتب عددها في العمود الثالث الذي يسمى عمود التكرارات ويقصد بها عدد الطلاب أمام كل تقدير. ولا بد وان يكون مجموعها يساوي مجموع التقديرات التفصيلية وهي في هذه الحالة 50 تقديراً.

وإذا حذفنا العمود الثاني من الجدول السابق (عمود العلامات) فإننا نحصل على جدول مألوف يتكون من عمودين ويسمى جدول التوزيع التكراري كما هو موضح في الجدول رقم (2) التالي:

الجدول رقم (2)

التوزيع التكراري لتقديرات توزيع الطلاب في مثال رقم (1)

التقدير	التكرار (عدد الطلاب)
A	12
B	20
C	7
D	7
F	4
المجموع	50

لابد من ملاحظة أن أي جدول إحصائي يحتوي على عنوان يوضح نوعية الجدول وطبيعة البيانات المعروضة فيه, كما يوضح مصدر هذه البيانات.

الجداول التكرارية الرقمية:

والجداول التكرارية الرقمية تستخدم في حالة قياس الظواهر بمقياس كمي (رقمي) مثل الأجور ودرجات الطلاب وقيم الإنتاج والاستهلاك وحسابات الدخل وتوزيع السكان ويرها. ولتوضيح طريقة تكوين الجدول التكراري الرقمي نأخذ المثال التالي:

مثال (2):

الجدول التالي يوضح عدد جوالات الذرة التي حصل عليها كل مزارع في 50 حواشة بمشروع الجزيرة للموسم الزراعي 90-1991م.

جدول رقم (3)

عدد جوالات الذرة التي حصل عليها كل مزارع

25	30	38	42	51	34	42	54	34	42
39	40	50	26	52	38	47	35	53	28
34	41	35	31	41	36	53	41	36	32
37	44	45	37	45	46	29	46	38	48
40	33	44	45	44	40	31	42	43	27

أولى الخطوات التي يجب اتخاذها لتنظيم هذه البيانات هي ترتيبها ترتيباً تصاعدياً أو تنازلياً وذلك لإعطاء فكرة عامة عن توزيع إنتاج الذرة. فإذا أخذنا بالترتيب التصاعدي للبيانات السابقة نتحصل على الجدول التالي:

جدول رقم (4)

الترتيب التصاعدي للبيانات في الجدول رقم (3)

54	50	46	44	42	40	38	35	33	29
53	48	45	44	41	40	37	35	32	28
53	47	45	43	41	39	37	34	31	27
52	47	45	42	41	38	36	34	31	26
51	46	44	42	40	38	36	34	30	25

ونلاحظ أن نصف الإنتاج يتراوح بين 44,35لحواشة الواحدة وان اعلى إنتاجية كانت 54 جوال واقل إنتاجية كانت 25 جوال. وعلى الرغم من أن فكرة الترتيب قد أبرزت لنا بعض المؤشرات إلا أنها ما زالت هنالك صعوبة في تفسير هذه البيانات. وتزداد هذه الصعوبة إذا كان المجتمع كبيراً.

ولتنظيم هذه البيانات وتلخيصها ووضعها في جدول تكراري نكون أولاً جدولاً للتفريغ كما فعلنا في حالة البيانات الوصفية مع استبدال الصفة في العمود الأول بمجموعات تسمى فئات.

وقبل ان نقوم بتكوين جدول التفريغ نستعرض فيما يلي طريقة اختيار الفئات.

أولاً تحديد عدد الفئات:

لتحديد عدد الفئات نتبع الخطوات التالية:

بداية نحدد مدى البيانات وهو يمثل الفرق بين اكبر قيمة في المفردات واصغر قيمة.وفي المثال رقم (2) يكون المدى كما يلي:

المدى (Range- R) = اكبر قيمة – اصغر قيمة

R = 54-25=29 (جوالاً)

ولا توجد قاعدة ثابتة لتحديد عدد الفئات ولكن عادة يقسم المدى إلى اعداد مناسبة من الفئات لا يقل عن 5 ولا يزيد عن 15 فئة تقريباً. كما يمكن الاستعانة بالقاعدة التالية لتحديد عدد الفئات

$$NC = 1 + 3.222 \log N$$
$$\therefore NC = 1 + 3.222 \log 50 = 6$$

حيث n هو عدد المفردات وفي هذا المثال نجد ان عدد الفئات يساوي 6 فئات تقريباً.

ثانياً : طول الفئة:

بتحديد المدى وعدد الفئات يمكننا بكل سهولة حساب طول الفئة من القانون التالي:

C = R/NC

حيث C تمثل طول الفئة و R هو المدى الكلى وn عدد الفئات وبالرجوع للمثال رقم (2) نجد ان طول الفئة هو 5.

C =29/6=5

ومن الأفضل دائماً أن يكون طول الفئة رقمياً دائرياً 5 او 10 أو 15...الخ

ثالثاً حدود الفئات:

بعد تحديد عدد وأطوال الفئات يصبح من الضروري ترسيم حدودها بحيث لا تتداخل مع بعضها البعض.وهنا من الأفضل دائماً أن نبدأ الحد الأدنى للفئة الأولى بأصغر عدد في المجموعة ما امكن ذلك. وعليه سيكون الحد الأدنى للفئة الأولى هو 25 ثم نضيف طول الفئة لنحصل على الحد الأعلى

للفئة الأولى وهو 30, ثم نبدأ الحد الأدنى للفئة الثانية وهـو يسـاوي الحـد الأعـلى للفئـة الاولى وهو 30 ثم نضيف طول الفئة لنحصل على الحد الأعلى للفئة الثانية... وهكذا.

باستخدام الخطوات السابقة يمكن تحديد فئات المثال رقم (2) على النحو التالي:

(30-25) , (35-30), (40-35), (45-40), (50-45), (55-50)

إذا كانت اطوال الفئات متساوية فيمكن كتابة الحد الادني فقط وتـرك الحـد الاعـلى الذي يتحدد تلقائياً بطول الفئة.

رابعاً: تكوين الجدول :

بعد الفراغ من تحديـد عـدد وأطـوال حـدود الفئـات نقـوم بتكـوين جـدول تفريـغ البيانات الكمية بنفس الطريقة التي اتبعناها عند تكوين جدول البيانات الوصـفية عـلى النحو التالي

جدول رقم (5)

جدول تفريغ البيانات لانتاج الذرة في مثال (2)

الفئات (إنتاج الذرة بالجوال)	العلامات	التكرار (عدد الحواشات)
25-	////	5
30-	/// ////	8
35-	//// ////	10
40-	/// //// ////	13
45-	/// ////	8
50-	/ ////	6
المجموع		**50**

لاحظ اننا اكتفينا بكتابة الحدود الدنيا للفئـات لان اطـوال الفئـات متسـاوية وهـي تقرأ كما يلى: من 25الى اقل من 30 , من 30 إلى اقل من 35... وهكذا

ويمكننا حذف عمود العلامات للحصول على جدول التوزيـع التكـراري مـن الجـدول رقـم (2-5), بحيـث يصـبح الجـدول مـن عمـودين فقـط: الاول يمثـل

فئات إنتاج الذرة بالجوال والثاني يمثل التكرارات لها أي عدد الحواشات المقابلة لكل فئة.

ويكتب كما يلي:

جدول رقم (6)

جدول التوزيع التكراري لانتاج الذرة في مثال رقم (2)

الفئات	التكرار f
25-	5
30-	8
35-	10
40-	13
45-	8
50	6
المجموع	50

خامساً: الحدود الفعلية للفئات:

ذكرنا سابقاً ان الفئة تتكون من حد ادنى وحد اعلى واللذان يسميان بحدود الفئات (Class Limits). نلاحظ ان حدود الفئات لها توزيع منقطع (Discrete) ولكن الحدود الفعلية للفئات يجب ان يكون لها توزيع متصل(Continuous) ولذلك نقوم باستبدال حدود الفئات بحدود متصلة تسمى حواجز الفئات(Class Boundaries) وتعرف في بعض الأحيان بالحدود الفعلية للفئات.

ويتم تكوين حواجز الفئات بطرح 0.5 من الحد الادنى للفئة لنحصل على الحاجز الادنى للفئة.

كذلك نضيف 0.5للحد الاعلى للفئة لنحصل على الحاجز الاعلى للفئة...وهكذا.

في المثال السابق يصبح التوزيع التكراري لحواجز الفئات كما يلي:

جدول رقم (7)

جدول التوزيع التكراري لانتاج الذرة

حواجز الفئات	التكرار f
29.5-24.5	5
34.5-29.5	8
39.5-34.5	10
44.5-39.5	13
49.5-44.5	8
54.5-45.5	6
المجموع	50

ونلاحظ اننا نرمز للتكرار f Frequency بالحرف

سادساً: مراكز الفئات:

تحسب مراكز الفئات والتي سنرمز لها بالحرف (X) من العلاقة التالية:

$$X = \frac{L+U}{2}$$

حيث L = الحاجز الادنى للفئة , U = الحاجز الاعلى للفئة. وإذا قمنا بتكوين مراكز الفئات للجدول رقم (7-2) نحصل على الجدول التالي:

جدول رقم (8)

جدول التوزيع التكراري لإنتاج الذرة في مثال (2)

حواجز الفئات	التكرار f	مراكز الفئات X
29.5-24.5	5	27
34.5-29.5	8	32
39.5-34.5	10	37
44.5-39.5	13	42
49.5-44.5	8	47
54.5-45.5	6	52
المجموع	50	

أنواع الجداول التكرارية:

ويسمى الجدول التكراري جدول منتظم اذا تساوت جميع فئاته. اما اذا اختلفت اطوال الفئات فيكون الجدول غير منتظم. ولا يلجأ الباحثون للجداول غير المنتظمة الا اذا كانت البيانات تتطلب ذلك كأن يتمركز عدد كبير من المفردات في مدى ضيق بينما ينتشر جزء كبير منها على مدى واسع. على سبيل المثال عند دراسة توزيع الدخل للافراد نجد على طرفي التوزيع اصحاب الدخول المنخفضة واصحاب الدخول العالية تنتشر على مدى واسع.

والجدول رقم (9) يعطي مثالاً للجدول التكراري غير المنتظم.

جدول رقم (9)

جدول التوزيع التكراري لمستويات الدخل الشهري

لعينة من الاسر (بالجنيهات السودانية)

التكرار f	فئات الدخل
500	600-
300	1000-
400	1500-
600	2000-
1000	2500-
1500	3000-
800	4000-
200	6000-
120	10000-
50	20000-
5450	المجموع

المصدر بيانات افترضية

من الواضح ان فئات الدخل في جدول رقم (9) لها اطوال متباينة.على سبيل المثال طول الفئة الاولى يساوي 400 بينما طول الفئة الثانية 500 والفئة قبل الاخيرة 10000, وبالتالي هو جدول تكراري غير منتظم.

الجداول التكرارية المفتوحة:

في بعض الدراسات يصعب تحديد الحد الادنى للفئة الاولى وفي هذه الحالة يسمى الجدول التكراري مفتوحاً من اسفل.كما يصعب احياناً تحديد الحد الاعلى للفئة الاخيرة فيسمى الجدول التكراري مفتوحاً من اعلى.ويمكن ان يكون الجدول مفتوح من الطرفين.

جدول رقم (10)

التوزيع العمري لعينة من النساء المتزوجات في المسح الصحي والديمغرافي السودان 89-1990م

فئات العمر	عدد النساء f
اقل من 15	380
-15	983
-20	1355
-25	970
-30	1047
-35	630
فاكثر40	540

المصدر وزارة التخطيط – مصلحة الاحصاء- المسح الصحي والديمغرافي ص 17- الخرطوم – مايو 1991

الجداول التكرارية المتجمعة:

في بعض الأحيان يتطلب البحث معرفة عدد أو نسبة المفردات التي تقل عن قيمة معينة أو تلك التي تزيد عن حد معين. يتم ذلك عن طريق الجداول التكرارية المتجمعة, وهي نوعان:

الجدول التكراري المتجمع الصاعد:

التكرار المتجمع الصاعد المقابل للحاجز الأعلى لفئة ما هو عدد القيم التي تقل عن ذلك الحاجز. أي هو مجموع تكرارت هذه الفئة والفئات التى قبلها. والجدول التكراري يسمى الجدول التكراري المتجمع الصاعد. بمعني آخر فإنه في الجدول التكراري المتجمع الصاعد يخصص العمود الاول للحواجز العليا للفئات والعمود الثاني للتكرار المتجمع الصاعد حيث يجمع تكرار كل فئة إلى مجموعة التكرارات السابقة لها. ولتوضيح هذا نوجد الجدول التكرارى المتجمع الصاعد من جدول رقم (11) الذي يتضمن انتاج الذرة في مشروع الجزيرة.

جدول رقم (11)

الجدول التكراري المتجمع الصاعد لانتاج الذرة بمشروع الجزيرة

التكرار المتجمع الصاعد	فئات الانتاج
0	اقل من 24.5
5	اقل من29.5
13	اقل من 34.5
23	اقل من39.5
36	اقل من44.5
44	اقل من49.5
50	اقل من54.5

يلاحظ ان التكرار المتجمع الصاعد المقابل لأقل من الحاجز الادني للفئة الأولى يساوي صفر والذي لأقل من الحد الأعلى للفئة الأخيرة يساوي المجموع أي 50 طالباً.

الجدول التكراري المتجمع الهابط:

في بعض الاحيان قد يتطلب الامر الحصول على التوزيع التكراري المتجمع لجميع القيم الاكبر من او المساوية للحاجز الادنى للفئة. وهذا يسمى التوزيع التكراري المتجمع الهابط. وفيه تمثل فئات الانتاج الحواجز الدنيا للفئات والتكرار المتجمع الهابط يمثل تراكم التكرارات من اسفل إلى اعلى.

جدول رقم (12)

الجدول التكراري المتجمع الهابط لانتاج الذرة بمشروع الجزيرة

التكرار المتجمع الهابط	فئات الانتاج
50	24.5 فأكثر
45	29.5 فأكثر
37	34.5 فأكثر
27	39.5 فأكثر
14	44.5 فأكثر
6	49.5فأكثر
0	54.5 فأكثر

الجدول التكراري النسبي:

التكرار النسبي لفئة هو تكرار الفئة مقسوماً على التكرار الكلي لجميع الفئات ويعبر عنه كنسبة مئوية. ويتم اعداد الجداول التكرارية النسبية دائماً من الجداول التكرارية البسيطة, في الجدول التالي نوضح التكرارات النسبية لانتاج الذرة في مشروع الجزيرة.

جدول رقم (13)

الجدول التكراري النسبي لانتاج الذرة في مثال (2)

التكرار النسبي	التكرار f	فئات الانتاج
10	5	29.5-24.5
16	8	34.5-29.5
20	10	39.5-34.5
26	13	44.5-39.5
16	8	49.5-44.5
12	6	54.5-45.5
100	50	المجموع

العرض البياني للبيانات:

يؤدي العرض البياني إلى زيادة إيضاح البيانات ويمكن للقارئ أن يكون فكرة سريعة بمجرد النظرة السريعة الفاحصة للرسوم البيانية، ولكن يؤخذ عليها أنه لا يمكن الاستفادة منها إلا في إلقاء نظرة أو فكرة عن البيانات حيث أنها توضح قدر بسيط من المعلومات.

طرق عرض البيانات:

تنقسم الأشكال والرسوم البيانية إلى قسمين رئيسيين:

عرض البيانات للجداول البسيطة وعرض البيانات للجداول التكرارية:

أولاً: العرض البياني للجداول البسيطة:

(أ) الخط البياني Line Chart

ويستخدم لتوضيح نمط اتجاه ظاهرة ما خلال فترة زمنية محددة، ولرسم الخط البياني يوضع الزمن على المحور الأفقي والظاهرة تحت الدراسة على المحور الرأسي، ثم تحدد النقاط حسب الإحداثيات المعطاة ثم توصل بخطوط

مستقيمة على سبيل المثال يمكن تمثيل البيانات الخاصة بتطور التعليم الابتدائي والمتوسط في السودان للفترة 74/1980 والواردة في الجدول (14) بخط بياني انظر شكل (14)

جدول (14)

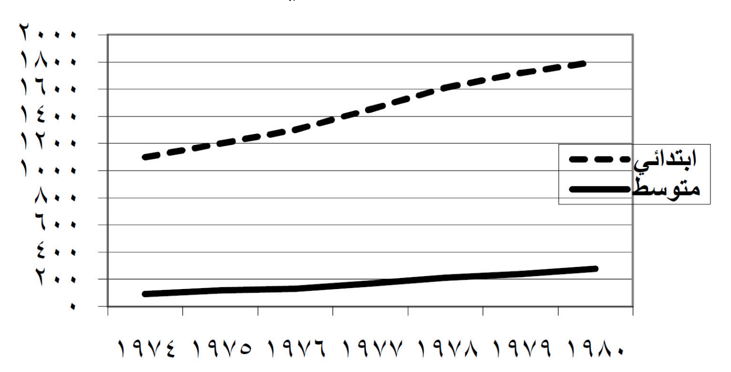
توزيع أعداد تلاميذ المدارس الابتدائية والمتوسطة في السودان خلال الفترة

العام	عدد التلاميذ (بالآلاف)	
	ابتدائي	متوسط
1974	1100	90
1975	1200	120
1976	1300	130
1977	150	170
1978	1610	210
1979	1720	240
1980	1800	279

المصدر الإحصاء التربوي – وكالة التخطيط التربوي 1985م.

شكل (1)

توزيع أعداد تلاميذ المدارس الابتدائية والمتوسطة في السودان خلال الفترة 1974- 1980



في بعض الأحيان تكون الظاهرتان تحت الدراسة مقاسة بنفس الوحدات ولها علاقة ببعضهما البعض، كما يكون للفرق بينهما معني معين، على

سبيل المثال الدخل و الإنفاق يعبر عنه بالادخار وهو الفرق بينهما، الصادرات والواردات يعبر عنه بالميزان التجاري، والمواليد والوفيات يعبر عنه بمعدل النمو السكاني وهكذا.

في مثل هذه الحالات تمثل كل ظاهرة بخط بياني كالمعتاد ويمثل الفرق بينهما المساحة المحصورة بين الخطين. تتضح هذه المسألة عند مقارنة المواليد و الوفيات بمحافظة الجزيرة في الفترة من 1980 – 1990م.

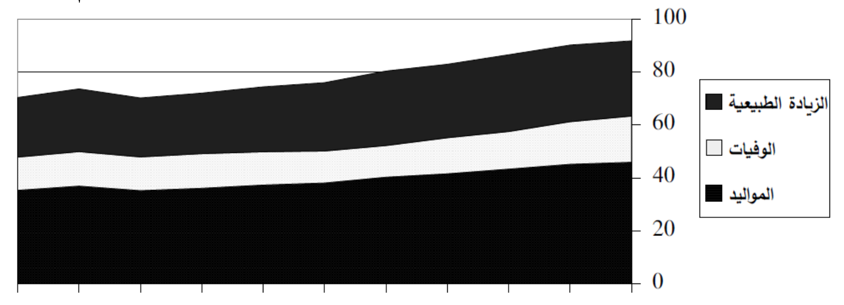
جدول (15)

المواليد والوفيات والزيادة الكلية محافظة الجزيرة 1980 -1990م (بالآلاف)

السنة	المواليد	الوفيات	الزيادة الطبيعية
1980	45.9	17.4	28.5
1981	45.9	16.1	29.0
1982	43.3	14.2	29.1
1983	41.5	13.6	27.9
1984	40.2	12.0	28.2
1985	38.0	12.1	25.9
1986	37.2	12.6	24.6
1987	6.0	13.1	22.9
1988	35.1	12.8	22.3
1989	36.8	13.1	23.7
1990	35.2	12.6	22.6

الشكل (2)

المواليد والوفيات والزيادة الطبيعية بمحافظة الجزيرة 1980 -1990م



ملحوظة:

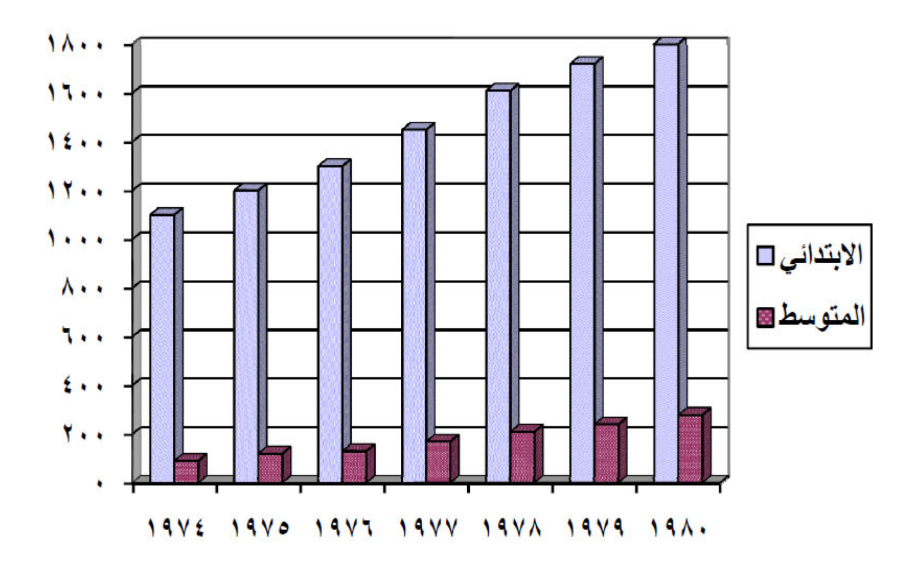
عند رسم الخط البياني لابد من تعيين مقياس رسم مناسب علـى كـل مـن المحـورين، فإذا كان مقياس الرسم على المحور الأفقي كبيرا مقارنة مع مقياس الرسم للمحور الـرأسي فإنه يمكن كسر المحاور وذلك سيؤدي إلى الزيادة في قيمة التغيرات الظاهرة عـلى الرسـم والعكس صحيحا، وتستخدم طريقة كسر المحاور كعلاج لمثل هذه الحالات.

(ب) الأعمدة البيانية: Bar chart

الأعمدة البيانية تمثل أعمدة رأسية قواعدها متساوية وتتناسب ارتفاعاتها مـع قـيم الظاهرة تحت الدراسة. المحور الأفقي في الأعمدة البيانية يمثـل الصـفة المميـزة للبيانـات والمحور الرأسي يمثل القيم المختلفة للظاهرة المعينة، ففي مثال تلاميذ المدارس المتوسـطة فإن تطـور أعـداد التلاميـذ خـلال الفـترة 1974 – 1976 يعـبر عنـه بـالتغير في ارتفاعـات الأعمدة كما هو موضح.

شكل(3)

تلاميذ المدارس المتوسطة في السودان خلال الفترة 1974 – 1980م:



ويلاحظ في الأعمدة البيانية عدم كسر المحور الأفقي، ولكن يجوز كسر الأعمـدة إذا كان الرسم يحتوي على عمود أو أكثر أطول بكثير من باقي الأعمدة. ويلاحظ في الأعمـدة البيانية عدم كسر المحور الأفقي، ولكن يجوز كسر الأعمدة إذا كـان الرسـم يحتـوي عـلى عمود أو أكثر أطوال بكثير من باقي الأعمدة.

وكسر العمود يمكن أن يكون جزئيا مع تكملته أفقيـا أو أن يـكسر الأعمـدة إذا كـان الرسم يحتوي على عمود أو أكثر أطول بكثير من باقي الأعمدة.

وكسر العمود يمكن أن يكون جزئيا مع تكملة أفقيـا أو أن يـكسر الجـزء الزائـد مـن العمود من أعلى مع كتابة القيمة العددية داخلة الجـزء المكسـور عـلى سـبيل المثـال إذا أخذنا الثلاث سنوات الأولى في جدول (3) وعبرنا عنها بأعمدة بيانية نجد أن أعداد طلاب المدارس الابتدائية أكبر بكثير من طلاب المدارس المتوسـطة وفي هـذه الحالـة يمكننـا كسر الأعمدة الخاصة بطلاب المدارس الابتدائية.

ملاحظات على الأعمدة البيانية:

1. لا يجوز كسر المحاور.
2. لابـد وأن تكـون قواعـد الأعمـدة متسـاوية وأن تكـون المسـافات بيـنهما أيضـا متساوية.
3. لا يجوز استخدام الأعمدة البيانية في حالة البيانات المتصلة.
4. يفضل استخدام النسب المئوية بدلا عن القيم الرقمية.

(ج) الرسوم الدائرية

وهي عبارة عن دائرة ذات نصف قطر مناسـب بحيـث يـتم تقسـيمها إلى قطاعـات تتلاقي في المركز وتتناسب مساحة هذه القطاعات مـع القـيم الجزئيـة لا ظهـار الأهميـة النسبية لكل مشاهدة، ويمكن حساب الزاوية المركزية للقطاع باستخدام القانون التالي:

$$\text{الزاوية المركزية للقطاع} = \frac{\text{عدد المشاهدات} \times 360}{\text{مجموع المشاهدات}}$$

ونوضح ذلك بالمثل التالي:

جدول (16)

يمثل وسائل منع الحمل المستخدمة لعينة من النساء بمركز ابن حيان بجزيـرة الفيـل لعام 1991

الوسيلة	عدد النساء	النسبة المئوية	الزاوية المركزية
الحبوب	60	60	216
اللولب	25	25	90
الحقن	15	15	54
المجموع	100	100	360

المصدر: مكتب الإحصاء – مركز ابن حيان – جزيرة الفيل 1991م

يلاحـظ أن مجمـوع النسـاء المـراد تمثـيلهم يسـاوي (1000) امـرأة قمنـا بتـوزيعهن بالنسبة المئوية كأول خطوة، وفي الخطوة الثانية حسبنا الزاوية المركزية لكل وسـيلة كـما يلي:

الزاوية المركزية للحبوب = 60/100 × 360 = 216

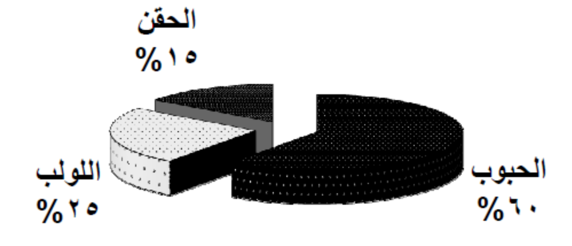
الزاوية المركزية للولب = 100/ 25 × 360 = 90

الزاوية المركزية للحقن = 100/ 15 × 360 = 54

الآن يمكن رسم الدائرة بعد تحديد نصف قطر مناسب وليكن في هذه الحالة 4 سـم كما يلي:

شكل (4)

استخدام وسائل منع الحمل بمركز ابن حيان جزيرة الفيل 1991م



والرسوم الدائرية تستخدم استخداما فعّالا في مقارنة البيانات، وهنا نرسم كل دائرة بحيث يتناسب نصف قطرها مع الجزر التربيعي لحجم الظاهرتين ويحسب ذلك من العلاقة التالية:

$$R_1/R_2 = \sqrt{X_1/X_2}$$

حيث أن X1, X2 تمثلان أجزاء المجاميع و R1, R2 تمثلان أنصاف أقطار كل من الدائرتين. وفيما يلي نقوم بعملية مقارنة بالدوائر على بيانات صادرات دولة الكويت عامي 1973 و 1983 حسب ما هو موضح في الجدول (17).

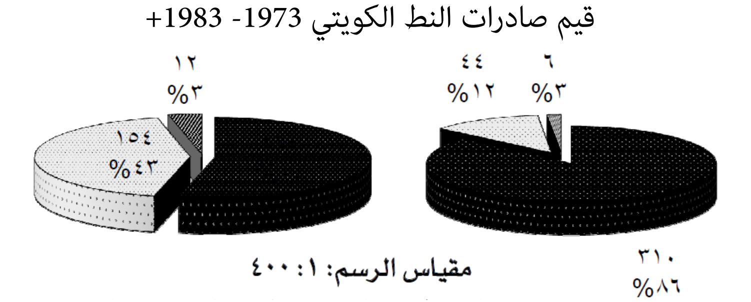
جدول (17)

قيم صادرات النفط لدولة الكويت 1973- 1983

	بيان 1973			بيان 1983		
	القيمة	%	الزاوية	القيمة	%	الزاوية
نفط خام	913022	86.2	310	1578171	53.7	194
منتجات نفطية	130870	12	44	1259269	42.9	154
غاز	16024	1.5	6	100752	3.4	12
المجموع	1059916	100	360	2938192	100	360
الجذر التربيعي				1714.1		
للمجموع	1029.5					
نصف القطر/سم	2.6			4.3		

شكل (5)

قيم صادرات النط الكويتي 1973- 1983+



المصدر: مصطفي الشلقاني: الإحصاء للعلوم الاجتماعية والتجارية 1989م

ثانياً: العرض البياني للتوزيعات التكرارية

(أ) المدرج التكراري Histogram

يستخدم المدرج التكراري لعرض بيانات الجدول التكراري وهو عبارة عن مستطيلات تمثل التكرارات المختلفة بحيث تتناسب مساحتها مع التكرارات وتتناسب أطوال قواعدها مع أطوال الفئات، وهنالك نوعان من المدرجات التكرارية

المدرج التكراري المنتظم:

في حالة الجدول التكراري ذو الفئات المتساوية تكون المستطيلات متساوية في العرض وتصبح النسبة بين مساحاتها كالنسبة بين تكراراتها.

مثال:

فيما يلي جدول تكراري يبين أوزان 100 طالب في أحدى الجامعات:

جدول (18)

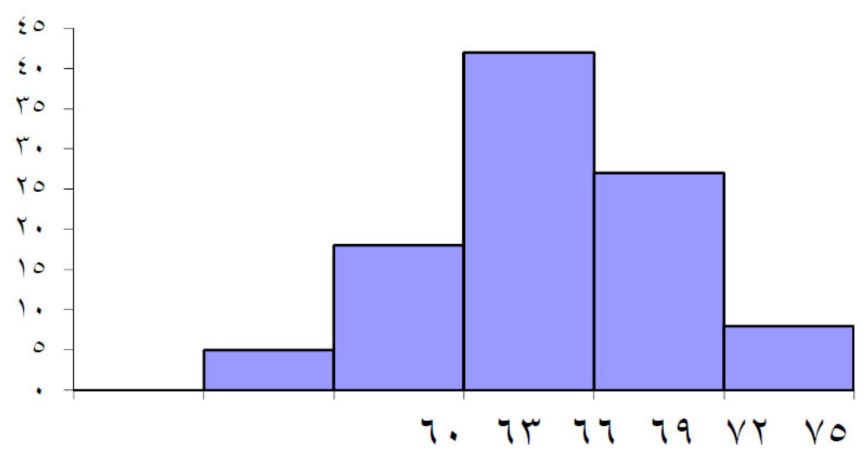
أوزان الطلاب في إحدى الجامعات

فئات الأوزان	عدد الطلاب f
-60	5
-63	18
-66	42
-69	27
75-72	8
المجموع	100

المصدر: مواري دي سبيجل – نظريات وسائل في الإحصاء – ماكجروهيل 1972

ولرسم المدرجة التكراري المنتظم يتم تقسيم المحور الأفقي إلى (5) أقسام متساوية تمثل بمقياس رسم مناسب بحيث يسمح بتمثل أكبر تكرارا في الجدول وهو 42 بعد ذلك نرسم خمس مستطيلات قواعدها متساوية وارتفاع كل منها يساوي تكرار الفئة المعنية ويمثل مركز الفئة منتصف قاعدة المستطيل.

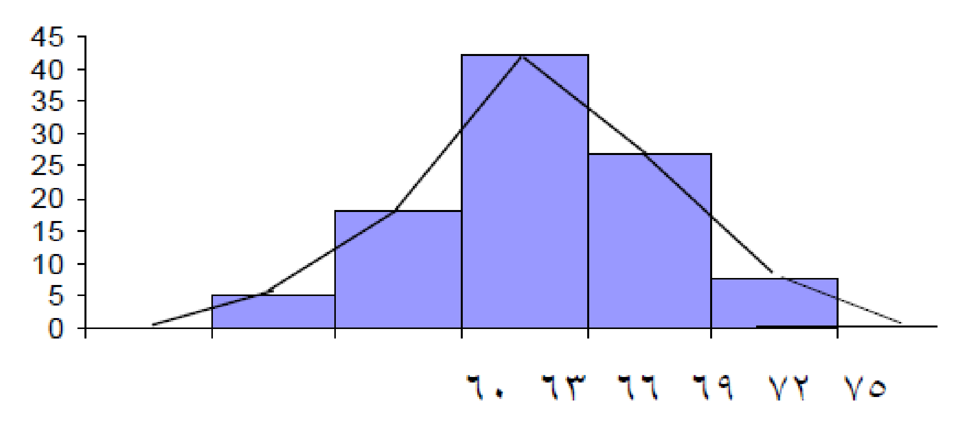

شكل (6) المدرج التكراري المنتظم لأوزان الطلاب في المثال أعلاه.



(ب) المضلع التكراري Frequency Polygon

وهو عبارة عن خط متقطع يمر بمنتصف المستطيلات (مراكز الفئات) من أعلى بعد إضافة وصلتين إلى ما بعد مركز الفئة الدنيا ومركز الفئة العليا كما في شكل (7).

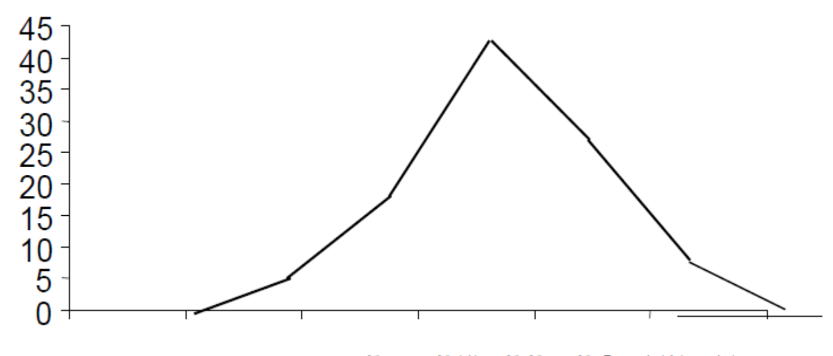
شكل رقم (7)المضلع التكراري لأوزان طلاب إحدى الجامعات



وفي المضلع التكراري فإن مساحة المثلثات الخارجية تساوي مساحة المثلثات الداخلية وعليه فإن مجموع مساحة المستطيلات تحت المدرج التكراري تساوي المساحة المحصورة بين المضلع التكراري والمحور الأفقي ويمكن رسم المضلع التكراري مباشرة دون رسم المدرج التكراري وذلك

بتحديد نقاط ارتفاعات المستطيلات مقابل مراكز الفئات ثم نوصل النقط بخطوط مستقيمة لنحصل على المضلع التكراري.


شكل (8) المضلع التكراري لأوزان طلاب إحدى الجامعات



(ج) المدرج التكراري غير المنتظم:

تظهر أهمية العلاقة بين مساحات المستطيلات التي تمثل الفئات وبين ارتفاعات الأعمدة التي تمثل تكرارات الفئات إذا كانت أطوال الفئات غير متساوية. ولتحقيق التناسب بين الأعمدة يتوجب تعديل التكرارات لنحصل على ما يسمى بالتكرار المعدل عن طريق قسمة التكرار الأصلي للفئة على طول الفئة. سوف نستخدم بيانات مستويات الدخل الشهري لعينة من الأسر الواردة في جدول (9) وقبل رسم المدرج التكراري غير المنتظم نقوم أولا بحساب التكرارات المعدلة في جدول (19) التكرارات المعدلة لمستويات الدخل الشهري لعينة من الأسر (بالجنيهات السودانية):

جدول (19)

التكرارات المعدلة لمستويات الدخل الشهري لعينة من الأسر (بالجنيهات السودانية):

فئات الدخل	التكرارات f	طول الفئة f	التكرار المعدل fc	fc × 100
-600	500	400	1.25	125
-1000	300	200	1.5	150
-1200	400	300	1.33	133
-2000	600	500	1.2	12
-2500	1000	500	2.0	200
-3000	1800	1000	1.8	180
-4000	800	2000	0.4	40

ملحوظة:

لقد قمنا بحذف فئات الدخل الثلاثة الأخيرة لتسهيل عملية الرسم كما تم ضرب التكرارات المعدلة (fc × 100) في رقم ثابت (100) للتخلص من الكسور شكل (9) يبين المدرج التكراري غير المنتظم للبيانات في جدول (19) ويتم الرسم بنفس الطريقة التي تم بها الرسم في حالة المدرج التكراري المنتظم وذلك بعد تعديل التكرارات.

والمضلع التكراري في التوزيعات غير المنتظمة يتم رسمه مباشرة عن طريق مراكز الفئات أو من المدرج التكراري، ولكن في الحالتين يجب أولا تعديل التكرارات قبل الرسم.

(ب) المنحني التكراري frequency curve

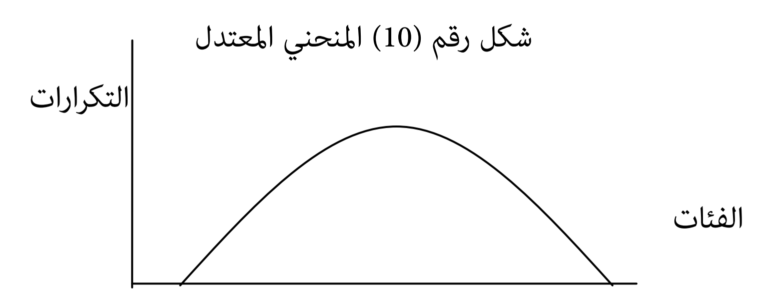
في علم الإحصاء يكون الغرض دائما من الرسومات والأشكال البيانية هو الوقوف على الاتجاه العام لها ومن ثم إيجاد المعادلة الرياضية التي خضع لها التوزيع التكراري، ولتكوين المنحني التكراري فأنه يتم تحديد نقاط المضلع التكراري ثم يرسم منحني يمر بهذه النقاط أو معظمها على أن يكون توصيل النقاط في شكل تمهيد باليد أو دقيقا باستخدام أحد الأساليب الرياضية.

والمنحنيات التكرارية نوعان:

المنحنيات المتماثلة في هذا النوع يتماثل المنحنى حول محور رأسي يمر بنقطة النهاية العظمى ويقسم المساحة أسفل المنحني إلى قسمين متطابقين كما في شكل (10).

وأكثر المنحنيات المتماثلة استخداما هو المنحنى المعتدل كما سنرى لاحقا وهو ناقوس الشكل وله قمة واحدة ويمتد طرفاه إلى ما لا نهاية.

شكل رقم (10) المنحني المعتدل



وفي الحياة العملية المنحنيات التكرارية المتماثلة قليلة الحدوث وجميع المشاهدات الإحصائية تؤول إلى شبه التماثل.

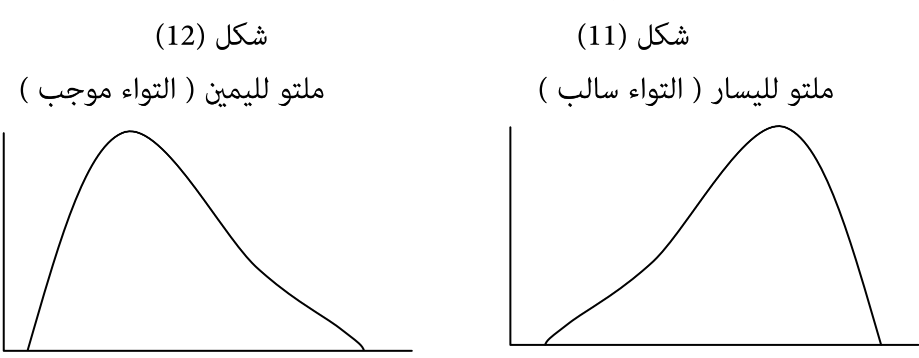
والمنحنيات غير المتماثلة: أما المنحنيات التكرارية غير المتماثلة فهي التي كون عدم تماثلها واضحا وهي تسمى منحنيات ملتوية، والمنحنى الملتوي يكون أحد طرفيه أطول من الطرف الآخر، فإذا زيل المنحنى إلى اليسار يكون الالتواء سالبا وإذا كان زيل المنحني إلى اليمين يكون الالتواء موجب. انظر الشكل (11) و (12)

شكل (11)

ملتو لليسار (التواء سالب)

شكل (12)

ملتو لليمين (التواء موجب)

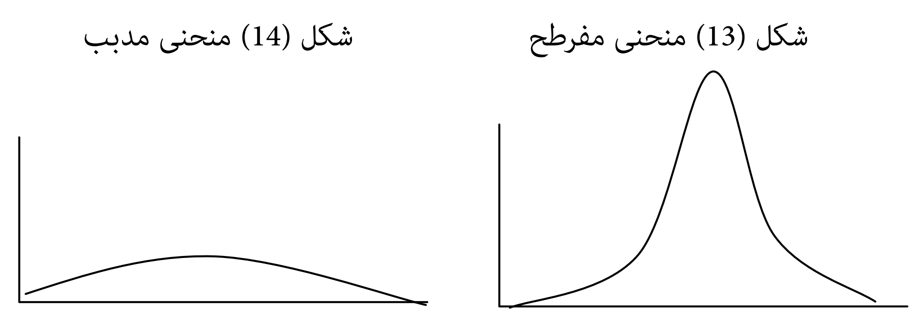


المنحنيات المفرطحة Kurtic curves

وهنا يكون الاختلاف من المنحنى المعتدل في شكل قمة المنحنى، فالمنحنى قد يكون مفرطحا إذا كان للتوزيع التكراري قمة حدباء بحيث أنها أعلى من قمة التوزيع المعتدل. انظر شكل (13) وشكل (14).

شكل (13) منحنى مفرطح

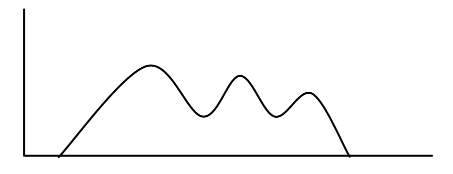
شكل (14) منحنى مدبب



المنحنيات متعددة القمم Multi- modal curve

عندما تكون مفردات المجتمع غير متجانسة ينتج عن ذلك منحنى تكراري متعدد القيم، وفي الغالب تظهر مثل هذه الحالة إذا كان المجتمع يتكون من عدة مجموعات مختلفة متداخلة انظر الشكل (15)

شكل (15) منحنى متعدد القمم

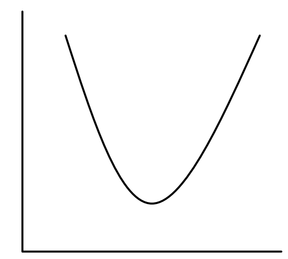


منحنيات أخرى:

وهنالـك أمثلـة أخـرى للمنحنيـات التكراريـة مثـل المنحنـى النـوني والمنحنـى الأسي والمنحنى الأسي المعكوس انظر أشكال (16) و (17)، (18).

شكل (16)

المنحني النوني



شكل (17)

المنحني الأسي



شكل (18)

المنحنى الأسي المعكوس



الفصل الرابع

مقاييس النزعة المركزية

Measures of Centeral Tendency

وهي تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها أغلبية البيانات وبحيث تمثلها أفضل تمثيل ومن مقاييس النزعة المركزية:

الوسط الحسابي the Arithematic Mean:

يعتبر الوسط الحسابي من أكثر مقاييس النزعة المركزية استعمالاً لكونه يستعمل جميع البيانات.

ويعرف الوسط الحسابي: مجموع هذه المشاهدات مقسوماً على عددها.

الوسط الحسابي للقيم غير المبوبة او البيانات الخام Raw Data

(X) الوسط الحسابي هو مجموع القيم مقسوماً على عددها.

الوسط الحسابي لمجموعة من القيم (n) (المشاهدات):

$$X_1 , X_2, X_3 , \dots, X_n$$

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots X_n}{n}$$

(n) للقيم من i = 1 ,2 , ...n (عدد القيم)

ويمكن استعمال Σ (سيغما) ويعني جمع الحدود التي في داخله.

ولذلك تكون معادلة الوسط الحسابي هي:

$$\bar{x} = \frac{\sum xi}{n} \Big/ \ \bar{x} = \sum_{f=1}^{n} xi/n$$

المثال:

أحسب الوسط الحسابي للبيانات التالية:

الحل:

$$\bar{X} = \frac{10+5+15+20+8+2}{6} = \frac{60}{6} = 10$$

مثال: أوجد الوسط الحسابي للمعادلة التالي:

$$\sum_{i=1}^{3}(2i + 2) \ n \Big/$$

الحل:

$$\frac{\sum_{i=1}^{3}(2i+2)}{n} = \frac{(2\times 1+2)(2\times 2+2)(2\times 3+2)}{3}$$

$$\bar{X} = \frac{4+6+8}{3} = \frac{18}{3} = 6$$

الوسط الحسابي من القيم المبوبة: (Grouped Data)

(التوزيع التكراري) فإن الوسط الحسابي يكون:

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{f_1 + f_2 + \cdots + f_n}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum f}$$

وحيث n = مجموع التكرارات.

مثال:

أوجد الوسط الحسابي للتوزيع التكراري الذي يمثل علامات 20 طالباً والعلامة القصوى هي 25 والتوزيع كمايلي:

العلامة	5	10	15	20	25
عدد الطلاب	8	3	2	3	4

الحل:

العلامة Xi	عدد الطلاب fi	Xifi
5	8	40
10	3	30
15	2	30
20	3	60
25	4	100
المجموع	20	260

$$\bar{X} = \frac{\sum xif}{\sum f} = \frac{260}{20} = 13$$

ولهذا فإن الوسط الحسابي لعلامات الطلبة هو 13.

مثال: جد الوسط الحسابي للتوزيع التكراري التالي:

الفئات	التكرار (fi)
10-14	8
15-19	15
20-24	10
25-29	9
30-34	5

الحل:

الفئات	التكرار fi	مراكز الفئات xi	Xifi
10-14	8	12	96
15-19	15	17	255
20-24	10	22	220
25-29	9	27	243
30-35	5	32	160
المجموع	47		974

$$\overline{X} = \frac{\sum xif}{\sum f} = \frac{974}{47} = 20.7$$

الوسط الحسابي المرجح (الموزون) The weighted mean

يفيد هذا المفهوم في حالات دمج مجموعات ذات أحجام عينات مختلفة إذا كان لدينا المجموعات التي أحجامها على التوالي:

$n_1 n_2,\ldots,n_r$

وأوساطها الحسابية على التوالي: x1,x2,...,xr

فإن الوسط الحسابي المرجح لهذه المجموعات يكون:

$$\overline{X} = \frac{(20\times45)+(30\times50)}{20+30}$$

$$= \frac{900+1500}{50}$$

$$= \frac{2400}{50} = 48$$

مثال: جد الوسط الحسابي المرجح للمجموعتان (x,y) حيث:

X = 16، 26، 8، 10، 15

y = 7، 16، 8، 10، 15، 4

الوسيط: Median

تعريف الوسيط:

هو القيمة التي يكون عدد القيم الأقل منها يساوي عدد القيم الأكبر منها.

وهو المقياس الثني من مقاييس النزعة المركزية في الأهمية. ويتم حسابه اذا تم ترتيب البيانات ترتيباً تصاعدياً أو تنازلياً.

الوسيط في حالة القيم المفردة والزوجية:

أ -الوسيط في حالة القيم المفردة:

اذا كانت القيم أو البيانات مفردة أي فردي نتبع الخطوات التالية لحساب الوسيط:

1-نرتب البيانات ترتيباً تصاعدياً أو تنازليا.

2-نجد رتبة الوسيط = $\frac{n+1}{2}$ حيث n عدد القيم.

3-الوسيط يساوي القيمة المقابلة للرتبة.

مثال: جد الوسيط للبيانات التالية: 16، 12، 10، 15، 8، 9، 6

الحل: ن عدد البيانات فردياً وبعد ترتيبها تصاعدياً تكون:

القيم 6 ،8 ،9 ،10 ،12 15 ،16

الرتب 1 ،2 ،3 ،4 ،5 ،6 ،7

نجد رتبة الوسيط = 4 = $\frac{7+1}{2}$

إذا رتبة الوسيط تساوي 4 فالوسيط يكون القيمة المقابلة للرتبة 4 ويكون (µ) = 10

ب -الوسيط في حالة القيم أو البيانات الزوجية:

اذا كان عدد القيم أو البيانات زوجياً نتبع نفس خطوات القيم أو البيانات المفردة ولكن يكون هناك رتبتان للوسيط لحساب الوسيط:

مثال: ما هو الوسيط للقيم التالية: 110، 90، 120، 80، 60، 100

الحل:

إن عدد البيانات زوجياً وبعد ترتيبها تصاعدياً تكون:

120، 110، 100، 90، 80، 60 القيم

6، 5، 4، 3، 2، 1 الرتب

نجد رتبة الوسيط = $\frac{n}{2}$ و $\frac{n}{2}+1$

$\frac{6}{2}=3$ و $\frac{6}{2}+1=4$

إذا القيمة المقابلة للرتبة 3 هي 90 والقيمة المقابلة للرتبة 4 هي 100 وبالتالي الوسيط يكون:

$$\mu = \frac{90+100}{2} = \frac{190}{2} = 95$$

الوسيط للبيانات المبوبة:

بما أن كثير من البيانات يتم عرضها بواسطة التوزيع التكراري فإن الفئة التي تقابل رتبة الوسيط والتي تساوي نصف مجموع التكرارات في جدول التوزيع التكراري المتجمع الصاعد، وتسمى فئة الوسيط (Medianclass) أو الفئة الوسيطية ويمكن إيجاد قيمة الوسيط باستعمال القانون التالي:

$$M = a + \left[\frac{\frac{\sum fi}{2} - f}{f}\right] \times C$$

$$M = a + \left[\frac{\frac{n}{2} - n1}{fm}\right] \times C$$

حيث: $\frac{n}{2}$ تمثل رتبة الوسيط، (n) مجموع التكرار.

A تمثل الحد الأدنى الفعلي لفئة الوسيط.

C: تمثل طول الفئة الوسيطية.

fm: تمثل تكرار الفئة الوسيطية.

n1: تمثل التكرار المتجمع الصاعد (السابق) الرتبة الوسيط.

n2: تمثل التكرار المتجمع الصاعد (اللاحق) لرتبة الوسيط.

$$fm = n_2 - n_1$$

مثال:

أوجد الوسيط لجدول التوزيع التكراري التالي والذي يمثل أعمار 50 شخصاً.

الفئات (الأعمار)	30-34	30-39	40-44	45-49	50-54
التكرار	7	10	4	12	17

الحل:

نضيف عمودين الأول الحدود الفعلية العليا والثاني التكرار المتجمع الصاعد ويصبح الجدول الجديد كمايلي:

الفئات	التكرار	الحدود الفعلية العليا	التكرار المتجمع الصاعد
30-34	7	34.5	7
35-39	10	39.5	17
40-44	4	44.5	21
45-49	12	49.5	33
50-54	17	54.5	50
المجموع	50		

ومن ثم نجد رتبة الوسيط:

$$\frac{n}{2} = \frac{50}{2} = 25$$

$$\frac{n}{2} = 25$$

وحيث أن رتبة الوسيط (25) تقع في التكرار المتجمع الصاعد أكبر مـن (21) وأصـغر من (33) فإذاً الحدود الفعلية لرتبـة الوسـيط هـي (49.5 – 44.5) وفي هـذه الحالـة (a) تكون (44.5) الحد الفعـلي الادني للفـة الوسـيطية. (c) تكـون (49.5 – 44.5) = 5 طـول الفئة الوسيطية (n1) تكون (21)، (n2) تكون (33)، لأن رتبة الوسـيط تقـع بـين هـذين الرقمين إذاً:

$$fm = 33 - 21 = 12$$

والآن نطبق قانون الوسيط:

$$\mu = 44.5 + \left[\frac{25-21}{12}\right] \times 5 \qquad = 44.5 + \frac{4}{12} \times 5$$

$$=44.5 + \frac{20}{12} \qquad =44.5 + 1.7 \qquad =46.2$$

مثال:

جد الوسيط لأوزان أمتعة (30) مسافر.

المسافرين التكرار	الأوزان الحدود الفعلية
5	10.5-15.5
6	15.5-20.5
7	20.5-25.5
8	25.5-30.5
4	30.5 -35.5

الحل:

التكرار	الحدود الفعلية	التكرار المتجمع الصاعد
5	10.5-15.5	5
6	15.5-20.5	(n_1) 11
7	20.5 -25.5	18 15رتبة الوسيط
8	25.5-30.5	(n_2) 26
4	30.5-35.5	30
30		

رتبة الوسيط = $\frac{n}{2} = \frac{30}{2} = 15$

والفئة الوسيطة (20.5-25.5) إذاً (a=20.5)

18-11 ،7=fm ،11=n1 ،5 = C

وبالتالي الوسيط يكون:

$$\mu = 20.5 + \frac{15-11}{7} \times 5 = 20.5 + 2.86 = 23.36$$

المنوال : The Mode

المقياس الثالث من مقاييس النزعة المركزية هو المنوال وهو القيمة الأكثر تكراراً بـين قيم المشاهدات. مع أحذ الملاحظات التالية بعين الاعتبار:

- يمكن أن يكون هناك منوال واحد.
- إذا لم يتكرر أي عدد نقول لا يوجد منوال.
- يمكن أن يكون هناك اكثر من منوال.
- إذا تكررت جميع القيم نفس عدد المرات فلا يوجد منوال.

المنوال من البيانات غير المبوبة:

هو القيمة الأكثر تكراراً بين القيم في المجموعة.

مثال: جد المنوال للبيانات التالية:

8 ،15 ،8 ،17 ،12 ،11 ،20 ،8 ،15 ،8

الحل: المنوال (Mode) هو رقم (8) (القيمة تكررت أكثر من غيرها).

مثال2: جد المنوال للبيانات التالية: 13، 16، 4، 5، 2، 8، 10

الحل: لا يوجد منوال (جميع القيم تكررت مرة واحدة).

مثال3: جد المنوال للبيانات التالية: 10، 11، 6، 10، 4، 3، 6

الحل: يوجد قيمتين للمنوال هما (6، 10) (يوجد أكثر من منوال).

مثال: جد المنوال للبيانات التالية: 6، 4، 2، 6، 4، 2

الحل: لا يوجد منوال (جميع البيانات تكررت نفس عدد المرات).

المنوال للبيانات المبوبة: The Mode of Grouped Data

المنـوال في التوزيـع التكـراري تكـون الفئـة التـي يقابلهـا أعـلى تكـرار تسـمى الفئـة المنوالية.

مثال: جد المنوال من الجدول التكراري التالي:

الراتب	100	150	220	400
التكرار (عدد العمال)	5	8	4	5

الحل: المنوال هو 150، القيمة المقابلة لأعلى تكرار وهو 8.

أما في التوزيع التكراري ذي الفئات فإن المنوال هو مركز الفئة المقابلة لأعلى تكرار.

مثال: جد المنوال التكراري التالي:

عدد الفئات	التكرار
10-18	4
19-27	6
28-36	3
37-45	2

الحل: المنوال هو مركز الفئة المقابل لأعلى تكرار وبما أن أعلى تكرار هو (6) فإن الفئة المنوالية هي (19-27) ومركزها يكون $\frac{19+27}{2} = 23$

إذاً المنوال يساوي (23)

كما يمكن إيجاد المنوال بطريقة الفروق لبيرسون وحسب الصيغة التالية:

$$\mu_0 = a + \left[\frac{d_1}{d_1+d_2}\right] \times C$$

مثال: أوجد المنوال في الجدول التكراري التالي:

الفئات	5-10	11-16	17-22	23-28
التكرار	4	8	12	2

الحل: تكون الجدول على الفئات الفعلية وكمايلي: ←ك

الحدود الفعلية	التكرار
4.5-10.5	4
10.5-16.5	8
16.5-22.5	12 الفئة المنوالية المقابلة لاعلى تكرار
22.5-28.5	2

الفئة المنوالية وهي (22.5 – 16.5) وحيث (a) تساوي الحد الأدنى الفئة المنوالية فإذا (a=16.5).

نجد (First difference) (d1) الفرق الأول = (4) = 8 – 12.

نجد (Second difference) (d2) الفرق الثاني = (10) = 2 – 12

ثم نجد المنوال حسب الصيغة المذكورة سابقاً:

$$\text{Mode} = 16.5 + \frac{4}{4+10} \times 6$$

$$= 16.5 + \frac{4}{14} \times 6$$

$$= 16.5 + \frac{24}{14}$$

$$= 16.5 + 1.7$$

$$= 18.2$$

خصائص الوسط الحسابي والوسيط والمنوال:

خصائص الوسط الحسابي:

1- سهل الحساب وأكثر مقاييس النزعة المركزية استعمالاً.

2- ياخذ بعين الاعتبار جميع قيم المشاهدات ذات العلاقة.

$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

3- مجموع انحرافات القيم عن وسطها الحسابي يساوي صفراً

$$\sum(x_1 - \bar{x}) = 0$$

مثال: إذا كانت لدينا البيانات التالية (7، 3، 2) أثبت أن مجموع انحرافات البيانات (القيم) عن وسطها الحسابي يساوي صفراً.

الحل: نجد الوسط الحسابي

$$\bar{X} = \frac{2+3+7}{3} = 4$$

$$= (2-4) + (3-4) + (4-7) = 0$$

$$= (-2) + (-1) + (3) = 0$$

4- يتأثر بالقيم المتطرفة:

مثال: أوجد الوسط الحسابي للبيانات التالية (1200، 20، 10، 30)

الحل: الوسط الحسابي:

$$\bar{X} = \frac{30+10+20+1200}{4} = \frac{1260}{4} = 315$$

وهذا العدد بعيد جداً عن باقي قيم المشاهدات وهذا بسبب القيمة المتطرفة 1200.

5- عند إضافة أو طرح أو قسمة أو ضرب عدد ثابت إلى جميع المشاهدات فإننا كذلك نفعل بالنسبة للوسط الحسابي.

مثال: أوجد الوسط الحسابي للقيم التالية: (30، 20، 10)

الحل: الوسط الحسابي:

$$\bar{X} = \frac{10+20+30}{4} = 20$$

خصائص الوسيط:

1. الوسيط لا يتأثر بالقيم المتطرفة.

مثال: أوجد الوسيط لقيم المشاهدات التالية:

15 ،20 ،30 ،45 ،290

رتبة الوسيط تكون = $3 = \frac{5+1}{2} = \frac{n+1}{2}$

القيمة المقابلة للرتبة (3) إذاً الوسيط هو (30) ونلاحظ أن القيمة المتطرفة (290) لم تؤثر على قيمة الوسيط.

2. يتأثر بعدد القيم والمشاهدات.

إذا حذفنا القيمتان الرابعة والخامسة من المثال السابق أعلاه تصبح رتبة الوسيط

$$\frac{n+1}{2} = \frac{3+1}{2} = 2$$

وقيمة الوسيط تساوي (20).

3. يأخذ بعين الاعتبار موقع القيم وليس متوسطها.
4. يمكن إيجاده من الجداول المفتوحة.
5. يمكن ايجاده بيانياً.

خصائص المنوال:

1. يمكن إيجاده بسهولة لكن قليل الاستخدام.
2. لا يتأثر بالقيم المنطرفة.

مثال: أوجد المنوال لقيم المشاهدات التالية:

300، 10، 50، 10، 6، 2

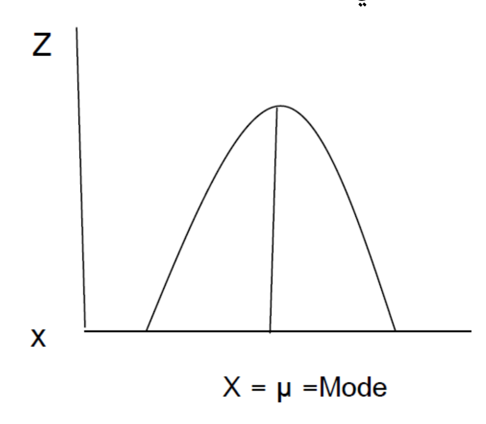
الحل: المنوال يساوي (10) القيمة الأكثر تكراراً وليتأثر بالقيمة المتطرفة (300).

3. يمكن إيجاده من الجداول المفتوحة.

4. يمكن إيجاده بيانياً.

العلاقة الخطية بين الوسط الحسابي والوسيط والمنوال:

- في التوزيعات وحيدة المنوال والمتماثلة فإن الوسط الحسابي والوسيط والمنوال تنطبق على بعضها البعض. مثل الشكل التالي:



- في التوزيعات التكرارية أحادية المنوال والملتوية إلتواء بسيط فإن الوسيط يقع في الوسط مابين الوسط الحسابي والمنوال. وتكون العلاقة بالصيغة التالية:

$$\bar{X} - mode = 3\,(x - \mu)$$

مثال: إذا كان الوسط الحسابي (30) والوسيط (28)، أوجد المنوال.

الحل:

$$30\text{-mode} = 3(30\text{-}28)$$

$$30\text{-mode} = 3(2)$$

30-mode = 6

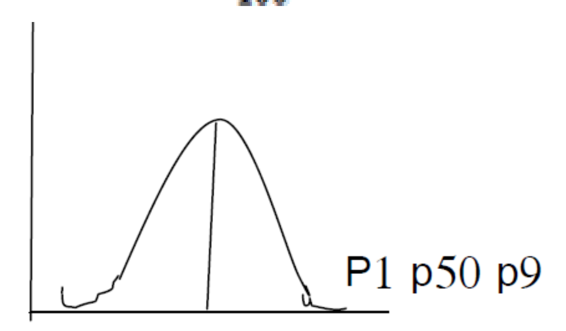
Mode = 30-6

Mode = 24

- في التوزيعات وحيدة المنوال والملتوية التواء بسيط نحو اليمين فإن العلاقة تكون المنوال أكبر من الوسيط أكبر من الوسط الحسابي.

الملينات والرتب الملينية Percentiles:

تحتاج في كثير من الاوقات لمعرفة نسبة البيانات التي تقع أو تزيد عن قيمة معينة أو تساويها. عندما تقسم المساحة أو البيانات إلى 100 جزء متساوية يسمى بالمنيات فالمنين الأول (P1) هو القيمة التي يسبقها 1% من البيانات ويليها 99% من البيانات وعلى فرض أن القيمة مرتبة ترتيباً تصاعدياً. والمتين (40) (P40) هو القيمة التي يسبقها 40% من البيانات ويليها 60% من البيانات، وهذه التقسيمات لها استخدامات كثيرة فمثلاً يمكن تقسيم مجموعة من العمال حسب الدخول ويمكن تحديد موقع كل عامل في هذه المجموعة. أما المئتين خمسون (P50) فهو القيمة التي تقسم البيانات إلى قسمين بحيث يكون نصف البيانات أقل من تلك القيمة أو تساويها، والنصف الآخر يكون أعلى من تلك القيمة أو تساويها. والشكل التالي يوضح مفهوم المنينات حيث المساحة تحت المنحنى وعلى كل جزء تساوي $\frac{1}{100}$ من المساحة الكلية.



علماً بأن المتين خمسون (P50) يساوي الوسيط يساوي الربع الثاني (Q2) ويساوي العشر الخامس (D5).

وبعد ترتيب البيانات ترتيباً تصاعدياً فإن المتين يرمز له بالرمز (pk) وحتى نستطيع التعرف على كيفية إيجاد المتين لابد من تفسير ذلك في حالة البيانات غير المبوبة والبيانات المبوبة.

المنينات من البيانات غير المبوبة:

إذا كانت البيانات غير مبوبة تتبع الخطوات التالية من أجل إيجاد المنينات:

- نرتب البيانات ترتيباً تصاعدياً.
- نجد ربة المنين من العلاقة التالية:

$$P_k = \frac{k}{100} \times n+1$$

- نجد موقع المئين ثم نجد قيمة المئين المناظرة لموقعه.

مثال: البيانات التالية تمثل أجور خمسة عمال في مصنع ما والمطلوب حساب المئين ثلاثون لهذه الأجور. (400، 300، 250، 350، 200).

الحل: نرتب الاجور ترتيباً تصاعدياً ونمنحها رتباً:

400، 350، 300، 250، 200 الأجور

5 4 3 2 1 رتب الأجور

ومن ثم نجد رتبة (p30):

$$P_{30} = \left[\frac{k}{100} \times n + 1\right]$$

$$P_{30} = \frac{30}{100} \times 5 + 1$$

وهي رتبة المئين 30 (p_{30}) 1.8 =

وتقع هذه الرتبة بين الرتبة (1) والرتبة (2).

نجد القيمة المناظرة للرتبتين الأولى والثانية وهما 200، 250. فتكون قيمة المئين ثلاثون هي:

$$P_{30}= \frac{200+250}{2} =225$$

وتفسير ذلك أن 30% من الأجور تقل عن 225 دولار وأن 70% من الأجور تزيد عن 225 دينار.

المئينات في البيانات المبوبة:

إن طريقة إيجاد المئينات في التوزيعات التكرارية هي نفس طريقة الوسيط، وذلك بإيجاد التوزيع التكراري المتجمع ثم فئة المئين وهي أول فئة يزيد تكرارها المتجمع على أو يساوي $\left[\frac{k}{100} \times n\right]$ حيث (n) تساوي مجموع التكرارات. ومن ثم نعين الحد الأدنى لفئة المئين (k) ونعبر عنه بالرمز (a) ونستعمل صيغة المعادلة التالية:

$$P_k = a + \frac{\frac{k}{100} \times n - n}{f} \times c$$

حيث:

a: تمثل الحد الأدنى فئة المئين.

$\left[\frac{k}{100} \times n\right]$: تمثل رتبة المئين.

nj: تمثل التكرار المتجمع السابق للفئة المئينية.

f: تمثل تكرار فئة المئين.

C: تمثل طول الفئة المئينية.

مثال: لدينا التوزيع التكراري التالي والذي يمثل أعمار شخصاً والمطلوب حساب p20.

الفئات	6-10	11-15	16-20	21-25	26-30	31-35	36-40
التكرار	5	7	11	25	17	9	6

الحل:

نكون الجدول التالي:

الفئات	التكرارf	الحدود	التكرار المتجمع الصاعد
6-10	5	5.5-10.5	5
11-15	7	10.5-15.5	12
16-20	11	15.5-20.5	23
21-25	25	20.5-25.5	48
26-30	17	25.5-30.5	65
31-35	9	30.5-35.5	74
36-40	6	35.5-40.5	80
المجموع	80		

- نجد رتبة المئين عشرون $P20 = \frac{20}{100} \times 80 = 16$
- نجد فئة المئين عشرون (15.5-20.5) مقابلاً لتكرار الصاعد وتساوي (0.1) أكبر من الرتبة (16).

إذاً:

$$12 = n1 ، 16 = \frac{20}{100} \times 80 ، 15.5 = a$$

$$20.5 - 15.5 = 5 = C \quad 11 = f$$

نطيق القانون:

$$P_{20} = 15.5 + \frac{16-12}{11} \times 5$$

$$P_{20} = 15.5 + 1.8 = 17.3$$

وتفسير ذلك أن 20% من الأشخاص أعمارهم تقل عن 17.5 سنة وأن 80% من الأشخاص أعمارهم تزيد عن 17.3سنة.

إيجاد (P_{60}) نفس معطيات المثال السابق:

نجد رتبة المئين ستون = $\frac{6}{100} \times 80 = 48$.

إذاً الرتبة (48) جاءت مطابقة لأحد التكرارات المتجمعة الصاعدة وهو (48) في هذه الحالة فإن (p_{60}) يساوي نهاية الفئة المناظرة للتكرار المتجمع الصاعد (48) ويساوي (25).

وحتى نستطيع إيجاد (P_1) نجد رتبة المئين الأول $\frac{1}{100} \times 80 = 0.8$

وهذه الرتبة أقل من التكرار الموجود للفئة الأولى وهو (5) في هذه الحالة لا يمكن إيجاد (P_1) إلا إذا أضفنا فئة سابقة لأول فئة ويكون تكرارها صفراً ومن ثم نتبع نفس الخطوات السابقة.

الربعيات والعشريات: Quartiles and Deciles

هذه التقسيمات تفيد في دراسات عديدة مجالات أخرى فالربيعات تعني تقسيم المساحة إلى أربعة أجزاء متساوية أو تقسيم المساحة إلى عشرة أجزاء متساوية.

الربعيات: Quartiles

القيم التي تقسم توزيع البيانات إلى أربعة أجزاء متساوية تعرف على أنها الربيعيات.

- الربيع الأول: (الربع الأدنى) هو القيمة التي يسبقها ربع البيانات ويليها ثلاثة أرباع البيانات ويرمز له بالرمز (Q_1).
- الربيع الثاني: له نفس قيمة الوسيط p_{50} ، D_5 (العشر الخاص) وهو القيمة التي يسبقها نصف البيانات النصف الآخر ويرمز لها ويليها (Q_2).
- الربع الثالث: (الربع الاعلى) وهو القيمة التي يسبقها ثلاثة أرباع ويليها ربع البيانات ويرمز له (Q_3).

الربعيات في البيانات غير المبوبة:

مثال: البيانات التالية تمثل أجور تسعة موظفين وهي كمايلي:

180، 160، 170، 150، 220، 200، 250، 190، 240

المطلوب: حساب Q1 , Q2 , Q3

الحل: نرتب البيانات ترتيباً تصاعدياً ونعطيها رتباً:

250، 240، 220، 200، 190، 180، 170، 160، 150

9 8 7 6 5 4 3 2 1

لايجاد $p_{25}=Q_1 = 2.5 = 1+9\times\frac{25}{100}$

رتبة الربع الأول $Q_1 = 2.5$ أكبر من 2 وأصغر من 3 فالقيمة المقابلة للرتب هي :

$= \frac{160+170}{2} = 165$ دينار.

لايجاد الربع الثاني $p_{25}=Q_2$ رتبة $Q_2 = p_{50} = 5 = 10 \times \frac{50}{100}$

إذاً الربع الثاني $Q_2=$ القيمة المقابلة للرتبة 5 وهي 190.

الربع الثالث (الربع الأعلى) = $Q_3 = P_{75}$

إذاً الرتبة $Q_3= 7.5 = 10 \times \frac{75}{100}$

إذاً الرتبة 7.5 تقع مابين الرتبة السابعة والثامنة.

إذا القيمة المقابلة لها هي $230 = \frac{220+24}{2}$

ونلاحظ هنا أننا نستخدم نفس طريقة المئينات ولكن أولاً يجب تحويل الربيعيات المطلوبة إلى مئينات علماً بأن:

$$P_{75} = Q_3 , P_{50}=Q_2, P_{25}=Q_1$$

الربعيات في البيانات المبوبة:

مثال: إذا كان لديك التوزيع التكراري والذي يمثل علامات 60 طالباً والتوزيع كما يلي:

الفئات	22-26	27-31	32-36	37-41	42-46	47-51
التكرار	10	8	12	9	13	8

والمطلوب: حساب Q_1 , Q_2

الحل:

الفئات	التكرار	الحدود الفعلية	التكرار المتجمع الصاعد
22-26	10	22.5-5 :26	10
27-31	8	26.5-31.5	18
32-36	12	31.5-36.5	30
37-41	9	36.5-41.5	39
42-46	13	41.5-46.5	52
47-51	8	46.5-51.5	60
المجموع	60		

لايجاد Q_1 والذي يساوي P_{25}

نجد رتبة $Q_1 = P_{25} = 15$ $\frac{25}{100}\times 60$

الرتبة 15 تقع في التكرار المتجمع الصاعد بين 10، 18.

$C=5$، $f=8$، $a=26.5$ $n_1=10$

$$Q_1 = P_{23} = 26.5 + \frac{15-10}{8} \times 5$$

$$Q_1 = 26.5 + 3.125 = 29.6$$

لايجاد الربع الثاني = $Q_2 = P_{50}$

رتبة الربع الثاني:

$$\frac{50}{100} \times 60 = 30$$

أو : $\frac{n}{2} = \frac{6}{2} = 30$

وحيــث أن الرتبــة 30 تقــع مقابــل التكــرار الصــاعد 30 إذاً Q_2=36.5 العشــيرات :Deciles

وهــي القــيم التــي تقســم البيانــات المرتبــة إلى عشرة أقســام متســاوية وهــي القــيم: $D_1,D_2,\ldots D_9$، فالعشر الأدنى هو العشر الأول D_1، والعشر الأعلى هو العشر التاسع D_9.

حساب العشريات في البيانات غير المبوبة:

البيانات التالية تمثل علامات سبعة طلاب والمطلـوب حسـاب العشـير الـثلاث (D_3) والعشير السابع (D_7) والبيانات كمايلي:

10 , 15 , 20 , 12 , 13 , 15, 18, 20

نرتب البيانات ترتيباً تصاعدياً:

10 ،11 ،12 ،13 ،15 ،18 ،20

1 2 3 4 5 6 7

نجد رتبة العشير الثالث $P_{30} = D_3 \quad \frac{30}{100} \times 8 = 2.4 =$

والرتبة 2.4 تقع بين 2-3

فنأخذ القيم المقابلة لها:

$$D_3 = \frac{11+12}{2} = 11.5$$

ولايجاد العشير السابع $p_{70} = D_7$

نجد رتبة D_7:

$$\frac{70}{100} \times 8 = 5.6$$

والرتبة لـ D_7 تقع بين 5 و 6

ناخذ القيم المقابلة لها فيكون D_7:

$$D_7 = \frac{15+18}{2} = 16.5$$

العشريات في التوزيع التكراري:

الفئات	التكرار	الحدود الفعلية	التكرار المتجمع الصاعد
22-26	10	22.5-5 :26	10
27-31	8	26.5-31.5	18
32-36	12	31.5-36.5	30 36
37-41	9	36.5-41.5	39
42-46	13	41.5-46.5	52
47-51	8	46.5-51.5	60
المجموع	60		

المطلوب: حساب العشير السادس D_6

نجد رتبة $D_6 = P_{60} = \frac{60}{100} \times 60 = 36$

$a = 36.5,\ n1=30,\ f=9,\ c=5$

$$\therefore D_6 = P_{60} = 36.5 + \frac{36-30}{9} \times 5$$

$$= 36.5 + \frac{6}{9} \times 5$$

$$= 36.5 + \frac{30}{9}$$

$$= 36.5 + 3.33 = 39.83$$

تمارين الفصل الرابع

1-البيانات التالية تمثل أجور مجموعة من العمال في إحدة المصانع وهي كمايلي:

180, 420, 250,500,300,400

175,195,190,200,350,150

والمطلوب:

-حساب الوسيط لاجور العمال.

-حساب الوسيط.

2-الجدول التالي يمثل أوزان مجموعة من الأطفال.

فئات الأوزان	عدد الأطفال F
4-10	12
11-16	10
17-22	23
23-28	8

المطلوب:

أ-حساب الوسط الحسابي لأوزان الأطفال.

ب-حساب الوسيط.

ج-حساب المنوال

د-أوجد العلاقة الخطية بين الوسط الحسابي والوسيط والمنوال.

3-أوجد الوسط الحسابي المرجح لأجور عمال مصنعين وبياناتهما كما يلي:

المصنع الاول الوسط الحسابي للأجور 450 وعدد العمال 120

المصنع الثاني الوسط الحسابي للأجور 350 وعدد العمال 70.

4-إذا كانت لديك البيانات التالية:

(5، 7، 12، 13، 3) أثبت بأن مجموع انحرافاتها عن وسطها الحسابي يساوي صفراً.

5-إذا كان الوسط الحسابي لمجموعة من البيانات هو 15، وتم ضرب جميع القيم بالعدد 2 فما هو الوسط الحسابي الجديد بعد عملية الضرب.

6-أذكر خصائص المنوال.

7-إذا كان الوسط الحسابي لمجموعة من البيانات هو 60 والوسيط 56 فما حجم المنوال.

8-البيانات التالية تمثل علامات سبعة طلاب في مادة مبادئ الإحصاء والمطلوب حساب المئين 40 لهذه العلامات وهي كما يلي:

(95,67,72,55,90,60,80)

9-الجدول التالي يمثل أجور عمال لمصنع أدوية:

فئات الاجور	عدد العمال
199-150	15
249-200	10
299-250	8
349-300	5
399-350	16
449-400	2

المطلوب:

أ-حساب الوسيط

ب-حساب المئين 60.

ج-حساب $Q_3(P_{75})$

د-حساب $D_5(P_{50})$

هـ-تمثيل هذه البيانات بيانياً بالمضلع التكراري.

الفصل الخامس

مقاييس التشتت

Measures of Dispersion

مقدمة:

إن المقاييس السابقة ومنها مقاييس النزعة المركزية غير كافية في بعض الأوقات لتحديد صفات التوزيعات التكرارية والبيانات الإحصائية. عندما يكون هناك ظاهرتان متساويتان في مقاييس الموقع كالوسط الحسابي والوسيط إلا أن الظاهرتين مختلفتان، في هذه الحالات لابد من استعمال مقاييس أخرى تبين مدى إختلاف البيانات فيما بينها ومدى التفاوت والتغير بين مفرداتها وهل هي متقاربة من بعضها أم هي متباعدة وعليه تستطيع مقاييس التشتت الإجابة عن مثل هذه الاسئلة بمعنى آخر تستخدم مقاييس التشتت للحكم على دقة مقاييس النزعة المركزية ومدى الاعتماد عليه في وصف البيانات. فإذا كان لدينا مجموعتان من القيم، الأولى (7، 4، 5) والثانية (1، 3، 8) فإن الوسط الحسابي لكل مجموعة هو (4). فنلاحظ أن الوسط الحسابي يعتبر قيمة نموذجية للمجموعة الأولى ويمثل بدقة كل قيمة، بينما نجد في المجموعة الثانية أن وسطها الحسابي قيمة غير نموذجية ولا يمثل مفردات هذه المجموعة بدقة. لذلك فإن مقاييس التشتت هي التي تعطي وصفاً أفضل من المقاييس السابقة.

التشتت يعني تباعد القيم عن بعضها البعض:

ويمكن تعريف التشتت (الاختلاف) على أنه التباعد بين المشاهدات داخل المجموعة الواحدة. ومقاييس التشتت هي تلك المقاييس التي تقيس مدى التشتت للمشاهدات عن وسطها الحسابي، مما يعني كلما كان التشتت كبير كلما دل ذلك على أن الفروقات كبيرة بين المشاهدات وعدم تجانسها (غالباً عدم التجانس يكون عند جمع المشاهدات من مجتمعات مختلفة)، وعندما تكون مقاييس التشتت قليلة فهذا يعني أن الفروقات بين المشاهدات داخل المجموعة الواحدة قليلة وأنها متجانسة مع بعضها البعض.

ومن مقاييس التشتت المدى، الانحراف المعياري والتباين والانحراف المتوسط ومعامل الاختلاف.

المدى (Range):

المدي في البيانات هو الفرق بين أعلى قيمة وأصغر قيمة. ونلاحظ في هـذا التعريـف بأن المدى لا يعتمد على جميع البيانات بل على أكبرهـا وأصـغرها فقـط، مـما يقلـل مـن أهميته. فإذا كانت القيمتان متطرفتان (قيم شاذة) عنـد ذلـك يكـون المـدى كبـيراً بيـنما المفردات الأخرى للبيانات ليست متباعدة عن بعضها البعض.

مثال: (1): أوجد المدى للبيانات التالية:

40، 12، 18، 16، 24، 30، 60

الحل:

نجد أكبر قيمة وأصغر قيمة فيكون المدى (R)

$$R = 60 - 12 = 48$$

مثال (2):

أوجد المدى للبيانات التالية:

1000، 150، 130، 120، 100، 10

الحل:

نجد أكبر قيمة وأصغر قيمة:

$$R=1000-10=990$$

ولكن هنا نلاحظ القيم المتطرقة (الشـاذة) وهـي أكـبر قيمـة وأصـغر قيمـة وبمـا أن المدى يعتمد عليها ينصح بحذف القيم المتطرفة الصغرى والكبرى. فيصبح المدى الجديـد هو:

$$R = 150-100=50$$

المدى الربيعي (Interquartile Range):

الفرق بين الربيع الثالث (الأعلى) والربع الأول (الادنى) Q_3-Q_1

مثال (3):

أوجد المدى الربيعي للبيانات التالية:

280، 210، 250، 220، 180، 170، 200، 160، 150، 190، 260

الحل:

نرتب البيانات ترتيباً ترتيباً تصاعدياً ثم نوجد الربع الثالث والربع الاول كما يلي:

280،	260،	250،	220،	210،	200،	190،	180،	170،	160،	150
11	10	9	8	7	6	5	4	3	2	1

ونجد أن رتبة الربع الأول:

$$Q_1 = P_{25} = \frac{25}{100} \times 12=3$$

$$\text{Or } Q_1 = ¼ \times n+1 = 3$$

إذاً الرتبة 3 يقابلها القيمة Q1 = 170.

رتبة الربع الثالث (الأعلى) هي:

$$Q_1 = P_{٧5} = \frac{75}{100} \times 12=٩$$

$$\text{Or } Q_1 = \times ¾ \; 12 = 9$$

القيمة 250(value) .

إذاً المدى الربيعي:

$$Q_3\text{-}Q_1$$

$$250\text{-}170=80$$

وعليه نصف المدى الربيعي يكون المدى الربيع/2

إذا نصف المدى الربيعي:

$$\frac{250-170}{2} = 40$$

مثال (4):

إذا كان الربع الثالث لمجموعة من البيانات يساوي 89 والربع الأول يساوي 39 وأعلى قيمة هي 100 وأدناه 40.

المطلوب: جد المدى ونصف المدى الربيعي

الحل:

المدى = أعلى قيمة – أدنى قيمة

100-40=60

نصف المدى الربيعي:

$$= \frac{Q3-Q1}{2}$$

$$= \frac{89-39}{2} \, 25$$

وبنفس الأجراء نستطيع إيجاد التالي:

- المدى المئيني = المئين الاعلى – المئين الأدنى
- نصف المدى المئيني = $\frac{\text{المدى المئيني}}{2}$
- المدى العشيري = العشر التاسع (الأعلى) – العشير الأول (الادنى)

D9 – D1

- نصف المدى العشيري = $\frac{\text{المدى العشيري}}{2} = \frac{D_9 - D_1}{2}$

المدى في البيانات المبوبة:

نجد المدى من العلاقة التالية:

الحد الفعلي للفئة العليا – الحد الفعلي الأدنى للفئة الدنيا

أو:

المدى = مركز الفئة العليا – مركز الفئة الدنيا

المدى الربيعي ونصف المدى الربيعي:

نستخدم نفس الصيغ للقيم المبوبة مع الأخذ بعين الاعتبار طرق استخراجها:

- المدى الربيعي = Q3-Q1
- نصف المدى الربيعي = $\frac{Q3-Q1}{2}$

ولتوضيح ذلك نأخذ المثال التالي:

مثال (5):

البيانات التالية تمثل أعمار (60) شخصاً:

الفئات الاعمار	22-26	27-31	32-36	37-41	42-46	47-51
تكرار الأشخاص	10	8	12	9	13	8

المطلوب: جد كل من:

1-المدى المطلق.

2-نصف المدى الربيعي.

الحل:

نكون جدول جديد كما يلي:

الفئات	التكرار	الحدود الفعلية	التكرار المتجمع الصاعد
22-26	10	22.5-5 :26	10
27-31	8	26.5-31.5	18
32-36	12	31.5-36.5	30
37-41	9	36.5-41.5	39
42-46	13	41.5-46.5	52
47-51	8	46.5-51.5	60
المجموع	60		

1-المدى = الحد الفعلي للفئة العليا – الحد الفعلي الأدنى للفئة الأخيرة

51.5-21.5=30

أو عن طريق مراكز الفئات:

49-24=25

2-نصف المدى الربيعي = $\frac{Q3-Q1}{2}$

لهذا نجد المدى ومن ثم نصف المدى ونتبع الخطوات التالي:

نجد رتبة الربع الأول (الأدنى):

$$Q_1 = \frac{25}{100} \times 60 = 15$$

نحدد موقع الربع الأول في عمود التكرار الصاعد حيث يقع بين 10 و 18، نحدد الفئة الربيعية وهي: 26.5-31.5.

نحدد الحد الأدنى للفئة الربيعية (a) : =26.5

نجد الربع الأول من العلاقة التالية:

$$Q_1= a+\frac{\frac{k}{100}\times n-n_1}{fm}\times C$$

$$Q_1 = 26.5 + \frac{15-10}{8}\times 5$$

$$=26.5+\frac{25}{8}$$

$$=26.5+3.1= 29.6$$

والآن نجد الربع الأعلى بنفس الأسلوب:

رتبة الربع الأعلى:

$$Q_3 = P_{75} = \frac{75}{100}\times 60 = 45$$

نحدد موقع الربع الأعلى في عمود التكرار الصاعد حيث يقع بين 39 و 52. نحدد الفئة الربيعية وهي: 41.5-46.5، ثم نحدد الحد الأدنى للفئة الربيعية (a) وهي الربع الثالق (الأعلى) في نفس العلاقة أعلاه:

$$Q_3 = 41.5 + \frac{45-39}{52-39}\times 5$$

$$= 41.5 + \frac{6}{13}\times 5$$

$$= 41.5 + \frac{30}{13}$$

$$= 41.5 + 2.3 = 43.8$$

الآن نجد المدى الربيعي:

$$Q_3-Q_1$$

$$43.8- 29.6 =14.2$$

ونجد نصف المدى الربيعي:

$$\frac{Q_3-Q_1}{2}$$

$$\frac{14.2}{2}$$

التباين والانحراف المعياري: Variance and standard Deviation

التباين: مجموع تربيع انحرافات البيانات عن وسطها الحسابي مقسوم على عدد البيانات – 1.

ونرمز للتباين بالرمز (S^2) وتكون صيغة معادلته كمايلي:

$$S^2 = \frac{\sum(x_i - x)^2}{n-1}$$

وترمز للانحراف المعياري بالرمز (S):

$$S = \sqrt{S^2}$$

أو:

$$S = \sqrt{\frac{\sum(x_i - x)^2}{n-1}}$$

مثال (6):

أوجد الانحراف المعياري من البيانات غير المبوبة التالية:

11، 6، 14، 10، 19

الحل:

1-نجد الوسط الحسابي:

$$x = \frac{19+10+14+6+11}{5}$$

$$= \frac{60}{5} = 12$$

2-نجد أولاً التباين من خلال إيجاد فروق البيانات عن وسطها الحسابي ثم نربعها ونجمعها ونقسمها على عدد البيانات مطروح منه (n-1).

$$S^2 = \frac{(19-12)^2+(10-12)^2+(14-12)^2+(6-12)^2+(11-12)^2}{5-1}$$

$$= \frac{(7)^2+(-2)^2+(2)^2+(-6)^2+(-1)^2}{4}$$

$$= \frac{49+4+4+36+1}{4}$$

$S^2 = \frac{94}{4} = 23.5 =$ التباين

ونلاحظ أنه تم إزالة الإشارة السالبة بتربيع الانحرافات.

أما الانحراف المعياري $S = \sqrt{S^2}$ = Standard Deviation

$$S = \sqrt{23.5} = 4.84$$

ويمكن إيجاد التباين للبيانات غير المبوبة من الصيغة التالية:

$$S^2 = \frac{\sum xi^2 - n\bar{x}^2}{n-1}$$

$$= \frac{814 - 5\times(12)^2}{5-1}$$

$$= \frac{814 - 5\times 144}{4}$$

$$= \frac{814 - 720}{4}$$

$$= \frac{94}{4} = 23.5$$

$$S = \sqrt{S^2} = \sqrt{23.5} = 4.84$$

مثال(7):

أوجد الانحراف المعياري في القيم المبوبة من الجداول الذي يمثـل التوزيـع التكـراري لأوزان مائة مسافر.

حدود فئات الأوزان	10-15	16-21	22-27	28-33	34-39	40-45
تكرار المسافرين	20	18	16	14	12	20

الحل:

نستخدم صيغة المعادلة التالية:

$$S^2 = \frac{\sum (x_i - \bar{x})^2 . f)}{n-1} \quad S$$

أو:

$$S = \sqrt{S^2}$$

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 . f)}{n-1}}$$

وعليه نكون جدول جديد كما يلي:

الفئات	التكرار f	مراكز الفئة xi	xi.f	xi-x	$(xi-x)^2$	$(xi-x)^2$.f
10-15	20	12.5	250	-14.4	207.36	4147.2
16-21	18	18.5	333	-8.4	70.56	1270.08
22-27	16	24.5	392	-2.4	5.76	92.16
28-33	14	30.5	427	3.6	12.96	181.44
34-39	12	36.5	438	9.6	92.16	1105.92
40-45	20	42.5	850	15.6	243.36	4867.2
المجموع	100		2690			11664

حيث أوجدنا مراكز الفئات (xi) وضربتها بالتكرار المقابل لها لكي نحسب الوسط الحسابي ثم أوجدنا انحرافات مراكز الفئات عن وسطها الحسابي وبعد ذلك نربع هذه الانحرافات ونضرب كل منها بالتكرار المقابل لها كما يظهر في الجدول السابق ونقسمها على مجموع التكرار مطروح منه واحد.

نجد الوسط الحسابي:

$$x = \frac{2690}{100} = 26.9$$

نطبق قانون التباين:

$$S^2 = \frac{\sum(x_i - x)^2.f}{n-1}$$

$$S^2 = \frac{11664}{100-1}$$

$$= 117.818$$

ثم نجد الانحراف المعياري حسب الصيغة التالية:

$$S = \sqrt{S^2}$$

$$S = \sqrt{117.818}$$

$$=10.85$$

وهناك صيغة أخرى لحساب التباين والانحراف المعياري.

التباين للبيانات غير المبوبة:

$$S^2 = \frac{(xi^2 - n\bar{x}^2)}{n-1}$$

xi^2f	xi^2	xi.f	مراكز الفئات xi	التكرار f
3125	156.25	250	12.5	20
6160.5	342.25	333	18.5	18
9604	600.25	392	24.536.5	16
13023.5	930.25	427	30.5	14
15987	1332.25	438	36.5	12
36125	1806.25	850	42.5	20
84025		2690		100المجموع

الوسط الحسابي هو:

$$\bar{x} = \frac{2690}{100} = 26.9$$

نعوض القانون ونجد التباين:

$$S^2 = \frac{84025 - 100 \times 26.9^2}{100-1}$$

$$S2 = \frac{84025 - 100 \times 723.61}{99}$$

$$S2 = \frac{84025 - 72361}{99}$$

$$S2 = \frac{11664}{99} = 117.818$$

والانحراف المعياري:

$$S = \sqrt{S^2}$$

$$S = \sqrt{117.818}$$

$$S = 10.85$$

الانحراف المتوسط The Mean Deviation:

هو متوسط الانحرافات المطلقة عن وسطها الحسابي ونلاحظ بأن الاشارة السالبة تلغي بسبب القيمة المطلقة ويرمز لها بالرمز (M.D).

الانحراف المتوسط للبيانات الأولية (غير المبوبة):

نستخدم القانون التالي:

$$M.D = \frac{\sum |xi - \bar{x}|}{n}$$

حيث أن $|xi - \bar{x}|$ هي القيمة المطلقة. بمعنى قيمة موجبة للفروقات.

مثال:

أجد الانحراف المتوسط للبيانات التالية:

16، 14، 8، 10، 12

الحل:

نوجد أولاً الوسط الحسابي ($\bar{x}$).

$$\bar{x} = \frac{12+10+8+14+16}{5}$$

$$\bar{x} = \frac{60}{5} = 12$$

ثم نجد القيمة المطلقة للفروق كمايلي:

$$M.D = \frac{|12-12|+|10-12|+|8-12|+|14-12|+|16-12|}{5}$$

$$= \frac{0+2+4+2+4}{5} = \frac{12}{5}$$

الانحراف المتوسط M.D = 2.4

الانحراف المتوسط في البيانات المجمعة (المبوبة):

ويأخذ الصيغة التالية:

$$M.\bar{D} = \frac{\sum |xi - \bar{x}| . fi}{\sum fi}$$

مثال: أوجد الانحراف المتوسط للجدو التكراري التالي:

الفئات	التكرار f	مراكز الفئةXi	xifi	$\|xi-x\|$	$\|xi-x\|.\bar{f}$
4-8	2	6	12	12.62	25.24
9-13	3	11	33	7.62	22.86
14-18	5	16	2.62	2.62	13.1
19-23	6	21	126	2.38	14.28
24-28	3	26	78	7.38	22.14
29-33	2	31	62	12.38	24.76
المجموع	21		391		122.38

$$\bar{x}=\frac{\sum xifi}{\sum fi}$$

$$\bar{x}=\frac{391}{21}$$

ونطبق القانون:

$$M.D=\frac{122.38}{21}=5.83$$ تقريباً

معامل الاختلاف (التغير) coefficient of Variation:

في بعض الأحيان يكون الانحراف المعياري وحده لا يكفي لإعطاء صورة واضحة عن التشتت للمجموعة من البيانات. وعلى هذا في أحياناً أخرى يكون معامل الاختلاف ملائم لمنح المدارس صورة أفضل عن التشتت داخل مجموعة القيم أو البيانات. وبذلك فإن معامل الاختلاف أو التغير يعطي أفضل من الانحراف المعياري وغيره من مقاييس التشتت لمقارنة تشتت البيانات أو القيم بين عدة مجاميع منها.

ويستخدم معامل الاختلاف أو التغير لمقارنة التشتت داخل عدة مجاميع من البيانات وإن كانت وحدات القياس للمجاميع مختلفة.

وكلما كانت النسبة أقل كان التشتت أقل ويرمز له بالرمز (C.V) ويعطي بالصيغة التالية:

$$C.V = \frac{S}{\bar{x}} \times 100\%$$

مثال:

إذا كانت أجور عمال في مصنعين متشابهين كما يـلي المصـنع الأول الوسـط الحسـابي للاجـور هـو (250) والمصـنع الثـاني الوسـط الحســابي للأجـور (400) دولار والانحـراف المعياري (5) والمطلوب حساب معامل الاختلاف لمعرفة في أي من المصنعين تتوزع الأجور بشكل أكثر عدالة.

الحل:

نجد معامل الاختلاف للمصنع الأول:

$$C.V = \frac{2.5}{250} \times 100$$

$$= \frac{250}{250} = 1\%$$

ثم نجد معامل الاختلاف للمصنع الثاني:

$$C.V = \frac{5}{400} \times 100 = 1.25\%$$

بالمقارنة نجد أن التشتت للأجور في المصنع الاول أقل منه في المصنع الثاني وعلى هذا الأساس الأجور أكثر عدالة في المصنع الاول.

تمارين الفصل الخامس

1-وضح المقصود بمقاييس التشتت والمدى.

2-إذا كانت أصغر قيمة هي 23 وأكبر قيمة هي 83. جد المدى.

3-البيانات التالية متوفرة لدينا وهي كمايلي:

(15، 8، 20، 14، 28، 35، 40، 50)

المطلوب:

-حساب المدى.

-حساب المئين.

-حساب المدى الربيعي.

4-البيانات التالية تمثل أوزان مجموعة من الطلاب:

الفئات (الأوزن)	التكرار (عدد الطلاب)
59-50	4
69-60	16
79-70	10
89-80	5
99-90	2

المطلوب:

أ-حساب نصف المدى الربيعي.

ب-الوسط الحسابي.

ج-العشير السابع.

5-أوجد الانحراف المعياري في البيانات التالية:

(22، 12، 28، 20، 38).

6-أوجد الانحراف المتوسط في البيانات التالية:

(8، 7، 4، 5، 6)

7-الجدول التكراري التالي يمثل أعمال من يمتلكون رخص قيادة:

الفئات	التكرار
25-18	100
32-26	180
40-33	40
48-41	20

المطلوب:

أ-حساب الانحراف المعياري.

ب-حساب الانحراف المتوسط.

8-إذا كان الوسط الحسابي لاجور عمال مصنع زيوت هو (280) دولار والتباين (100) والوسط الحسابي لمصنع آخر هو (340) والانحراف المعياري (15).

المطلوب: حساب معامل التغير لمعرفة في أي من المصنعين تعتبر الأجور أكثر عدالة.

الفصل السادس
الألتواء والتفرطح والعزوم

مقاييس الالتواء:

لمقاييس النزعة المركزية أنه إذا كانت المنحنيات معتدلة (متماثلة) فإن مقاييس الوسط الحسابي والوسيط والمنوال تكون متطابقة، أي لها نفس القيم. أما في حالة عدم التماثل فإنها تختلف.

لنرمز للوسط الحسابي و الوسيط و المنوال بـ $\bar{x}, \tilde{x}, \hat{x}$ بالترتيب

1/ إذا كان المنحنى ملتوياً نحو اليسار فيكون الوسط الحسابي أصغرها وتنطبق العلاقة:

$$\bar{X} < \tilde{X} < \hat{X}$$

2/ إذا كان المنحنى ملتوياً نحو اليمين، فيكون الوسط الحسابي أكبرها وتنطبق العلاقة:

$$\bar{X} > \tilde{X} > \hat{X}$$

في الحالة الأولى يكون الالتواء سالباً بحيث يكون ذيل التوزيع الأطول نحو اليسار. أما في الحالة الثانية فيكون الالتواء موجباً بحيث يكون ذيل التوزيع الأطول نحو اليمين. ويتم تحديد نوعية الالتواء باستخدام الفرق بين الوسط والمنوال فإذا كان:

(أ) $\left(\bar{X} - \hat{X}\right)$ قيمة موجبة يكون الالتواء موجبا.

(ب) .$\left(\bar{X} - \hat{X}\right)$ قيمة سالبة يكون الالتواء سالبا

(ج) $\left(\bar{X} - \hat{X}\right)$ تساوي صفر يكون التوزيع معتدلا.

وتبين الحالات الثلاثة في شكل (1):

معامل الالتواء:

لقد تعرفنا فيما سبق على العلاقة بين الوسط الحسابي والوسيط والمنوال عندما يكون التوزيع التكراري شبه متماثل وأوضحنا أن العلاقة التالية تتحقق:

$$\left(\bar{X} - \hat{X}\right) = 3\left(\bar{X} - \tilde{X}\right)$$

وقد استخدم كارل بيرسون هذه العلاقة في اشتقاق مقياسين للالتواء B_1 , B_2 على النحو التالي:

$$B_1 = \left(\bar{X} - \hat{X}\right)/\sigma_X$$
$$B_2 = 3\left(\bar{X} - \tilde{X}\right)/\sigma_X$$

ويسمي المعامل معامل B_1 بيرسون الأول B_2 معامل بيرسون الثاني للالتواء.

مثال:

في توزيع أجور 65 عاملا في إحدى الشركات (الوحدة الثانية) كان:

الوسط الحسابي $\bar{X} = 79.76$ و الوسيط $\tilde{X} = 79.06$ والمنوال $\hat{X} = 77.5$

وكان الانحراف المعياري $\sigma = 15.6$. أوجد معاملي بيرسون الأول والثاني للالتواء.

لإيجاد معاملي بيرسون نقوم بالتعويض المباشر في المعادلتين السابقتين:

* معاملا بيرسون الأول والثاني للالتواء:

$$B_1 = \frac{79.76 - 77.50}{15.60} = 0.14$$

$$B_2 = \frac{3(79.76 - 79.06)}{15.60} = 0.13$$

بما أن B_1 , B_2 موجبة فإن التوزيع ملتوي التواء موجبا أي أنه ملتوي ناحية اليمين ولكن من الواضح أن درجة الالتواء خفيفة للغاية وذلك لان:

B1 = 0.14 , B2 = 0.13

مما يعني أن لهم قيم متقاربة وعليه فإن العلاقة التقريبية بين الوسط والوسيط والمنوال قد تحققت الأمر الذي يدل على قرب التوزيع من التماثل. ويمكن قياس الالتواء من الانحراف الربيعي عن طريق معامل الالتواء الربيعي

B. $$B = \left(R_3 - 2R_2 + R_1\right)/\left(R_3 - R_1\right)$$

يمكن حساب معامل المدى الربيعي من بيانات إنتاج الذرة في مثال رقم 2 الفصل الثاني حيث وجدنا أن:

$$R_1 = 34.19$$

$$R_2 = 40.27$$

$$R_3 = 45.44$$

وعليه فإن معامل الالتواء الربيعي:

$$B = (45.44 - 2 \times 40.27 + 34.19)/(45.44 - 34.19) = 0.08$$

والأساس في هذا المعامل هو أنه في حالة التوزيعات المتماثلة تقع الربيعات الأول والثالث على بعدين متساويين من الوسيط، في حين يختلف بعداهما عنه في حالة التوزيعات الملتوية أي أنه عندما يكون قيم B أقرب إلى الواحد الصحيح.

مزايا وعيوب معامل الالتواء:

من مزايا هذا المعامل انه يمكن إيجاده بالحساب أو بالرسم باستخدام المنحنى التكراري المتجمع الصاعد. وكذلك يمكن حسابه في حالة الجداول المفتوحة.

ولكن عيبه الأساسي أنه لا يخذ جميع القيم في الاعتبار، بل يعتمد على القيم الواقعة بين الربيعين الأول والثالث فقط. كما أنه لا يمكن استخدامه في التحليل الإحصائي الرياضي.

مقاييس التفرطح:

ونستعمل كلمة تفرطح لوصف تدبب المنحني التكراري بالنسبة للمنحني المعتدل فإذا كان المنحني أكثر تفرطحا من المنحني المعتدل يوصف بأنه مفرطح وهو الذي يتسع في الوسط وتنحني قمته عن قمة المنحني المعتدل. إذا كان المنحني أقل تفرطحا من المنحني المعتدل يوصف بأنه منحني مدبب وهو الذي يأخذ مدى ضيقا في وسطه وترتفع قمته عن قمة المنحني المعتدل. أما إذا كان المنحني التكراري له قمة مطابقة للمنحني المعتدل فاه يوصف بأن تفرطحه متوسط (Mesokartic) أنظر شكل (2).

وتستخدم العزوم لقياس معاملي الالتواء والتفرطح. لذلك فأننا سنقوم بشرح مفهوم العزوم وطريقة حسابها:

العزوم Moments :

إذا كان لدينا مجموعة من المفردات $X_1 , X_2 \ldots, X_N$ فان العزم الـرائي حـول الصـفر ويرمز له بالرمز (X_r) يحسب من العلاقة التالية:

$$\bar{X}_r = \frac{\sum X^r}{n}$$

وعلى وجه الخصوص عندما تكون r = 1 فان العزم الأول حـول الصـفر هـو الوسـط الحسابي $\bar{X}$

كذلك العزم الرائي حول الوسط الحسابي ويرمز إليه بالرمز(m_r) يحسب مـن العلاقـة التالية:

$$m_r = \sum \left(X - \bar{X}\right)^r / n$$

فإذا كانت r = 1 فإن العـزم الأول حـول الوسـط الحسـابي = 0 أمـا إذا كانـت r = 2 فإن العزم الثاني حول الوسط الحسابي يساوي التباين σ^2.

وهنالك أيضا العزم الرائي حول اى نقطة A ويرمز لـه بـالرمز $\left(m'_r\right)$ ويحسـب مـن العلاقة التالية:

$$m'_r = \sum \left(X - A\right)^r / n$$

ونلاحظ انه إذا كانت $A = \bar{X}$ فان العزم الرائي حول النقطة A يساوي العزم الـرائي حول الوسط الحسابي. وكذلك إذا كانت A = 0 فان العزم الرائي حول النقطـة A يسـاوي العزم الرائي حول الصفر. وعليه فان m'_r أكثر عمومية وتعتمد قيمته عـلى نقطـة الأصـل A.

ونلاحظ أيضا أن $d_1 = x_1 - A$ هي انحرافـات القـيم عـن اى وسـط فـرضي A والتـي استخدمناها في حساب الوسط الحسابي والانحراف المعياري في الأقسام السابقة.

ويستخدم العزم الرائي حول النقطة A في اشتقاق علاقات مختلفة بين العزوم وهى:

$$\begin{aligned} m_1 &= 0 \\ m_2 &= m'_2 - m'_1 \\ m_3 &= m'_3 - 3m'_1 m'_2 + 2m'^3_1 \\ m_4 &= m'_4 - 4m'_1 m'_3 + 6m'^2_1 m'_2 - 3m'^4_1 \end{aligned}$$

ونقوم بإثبات m_2 فيما يلي:

$$m_2 = \sum\left(X - \overline{X}\right)^2 / n$$

$$m_2 = \left[\sum X^2 / n - \left(\sum X / n\right)^2\right]$$

$$d = X - A$$

$$X = A + d$$

$$\overline{X} = A + \overline{d}$$

$$\left(X - \overline{X}\right) = \left(d - \overline{d}\right)$$

$$\sum\left(X - \overline{X}\right)^2 / n = \sum\left(d - \overline{d}\right)^2 / n$$

$$= \left\{\sum d^2 / n - \left(\sum d / n\right)^2\right\}$$

$$= \left\{\sum (X - A)^2 / n - \left(\sum X - A / n\right)^2\right\}$$

$$\therefore m_2 = m'_2 - m'^2_1$$

مثال:

أوجد العزوم الثلاثة الأولى حول الـرقم 4 لمجموعـة المفـردات التاليـة 10، 8، 7، 3، 2 ومن ثم أوجد العزم الثاني والثالث حول الوسط الحسابي:

$$m'_1 = \sum (x - 4)/n = \frac{(2-4) + (3-4) + (7-4) + (8-4) + (10-4)}{5} = 2$$

$$m'_2 = \sum (x-4)^2 / n = \frac{(2-4)^2 + (3-4)^2 + (7-4)^2 + (8-4)^2 + (10-4)^2}{5} = 13.2$$

$$m'_3 = \sum (x-4)^3 / n = \frac{(2-4)^3 + (3-4)^3 + (7-4)^3 + (8-4)^3 + (10-4)^3}{5} = 59.6$$

ثم تستخدم العلاقة بين العزوم لحساب العزم الثاني والثالث حـول الوسـط الحسـابي كما يلي:

$$m_2 = m'_2 - m'^2_2 = 13.2 - (2)^2 = 9.2$$

$$m_3 = m'_3 - 3m'_1 m'_2 + 2m'^3_1$$

$$59.6 + 3(2)(13.2) + 2(2)^3 = 3.6$$

حساب العزوم من البيانات المبوبة:

إذا كانت $x_1 , x_2 , \ldots, x_n$ هي مراكز فئات لها تكـرارات $f_1 , f_2 , \ldots ,f_n$ فـان العـزوم السابقة يمكن أن تكتب بتكرارات كما يلي:

$$\overline{X}_r = \sum fx^r / n$$

$$m_r = \sum f\left(X - \overline{X}\right)^r / n$$

$$m'_r = \sum f\left(X - A\right)^r / n$$

وبما أن هذه الصيغ غير ملائمة للحسـاب نسـبة لطـول العمليـة الحسـابية ولتسـهيل الأمر سنلجأ إلى طرق الترميز السابقة وذلك لاستخدامها بصورة مخـتصرة لإيجـاد العـزوم. ولتوضيح هذه الطرق أكثر دعنا نتوصل إليها بالخطوات التالية:

ضع $d = x - A$ عليه فإن: $m'_r = \sum fd^r / \sum f$

والآن ضع $u = d/c$ فتصبح: $m'_r = c^r \left(\sum fu^r / \sum f\right)$ ويمكن اسـتخدام هـذه الصـيغ للحصول على mr بتطبيق معادلة اشتقاق العزوم.

مثال (4): أوجد العزوم الأربعة حول الوسط الحسـابي لتوزيـع إنتـاج الـذرة بمشروع الجزيرة.

يتم أولا تنظيم العمل وفق الجدول التالي:

جدول رقم (1)

حساب العزوم من بيانات إنتاج الذرة بمشروع الجزيرة

فئات الإنتاج	f	x	d=x-A	u=d/c	fu	fu^2	fu^3	fu^4
-27.5	5	27	-15	-3	-15	45	-132	405
-29.5	8	32	-10	-2	-16	32	-64	128
-34.5	10	37	-5	-1	-10	10	-10	10
-39.5	13	42	0	0	0	0	0	0
-44.5	8	47	5	1	8	8	8	8
-49.5	6	52	10	2	12	24	48	96
المجموع	50				-21	119	-153	647

$$m_1' = c\left(\sum fu/\sum f\right) = (5)(-21/50) = -2.1$$
$$m_2' = c^2\left(\sum fu^2/\sum f\right) = (5)^2(119/50) = 59.5$$
$$m_3' = c^3\left(\sum fu^3/\sum f\right) = (5)^3(153/50) = -382.5$$
$$m_4' = c^4\left(\sum fu^4/\sum f\right) = (5)^4(647/50) = 8087.5$$

الان نحسب العزوم الرابعة الاولي حول الوسط الحسابي بالاستفادة من هذة النتـائج (حسب التعريف):

$$m_1 = 0$$
$$m_2 = m_2' - m_1'^2 = 59.5 - (2.1)^2$$
$$m_2 = 54.1$$
$$m_3 = m_3' - 3m_1'm_2' + 2m_1'^3 = 382.5 + (3)(-2.1)(59.5) + 2(2.1)^3$$
$$m_3 = -26.1$$
$$m_4 = m_4' - 4m_1'm_3' + 6m_1'^2 m_2' - 3m_1'^4$$
$$= 8087.5 - (4)(2.1)(-382.5) + (6)(-2.1)^2(3)(-2.1)^4 = 12902.65$$
$$m_4 = 12902.65$$

حساب معامل الالتواء باستخدام العزوم:

حساب معامل الالتواء باستخدام العزم الثالث حول الوسـط الحسـابي والعـزم الثـاني حول الوسط الحسابي ويرمز إليه بالرمز $\alpha 3$، ويعرف كما يلي:

$$\alpha_3 = \frac{m_3}{\sqrt{(m_2)^3}}$$

مثال:

أوجد معامل الالتواء باستخدام العزوم لبيانات إنتاج الذرة بمشروع الجزيرة.

الحل:

كنا قد أوجدنا العزوم الأربعة حول الوسط الحسابي لهذا التوزيع ومـا علينـا الآن إلا التعويض في القانون السابق لإيجاد ($\alpha 3$) معامل الالتواء.

$$\alpha_3 = \frac{m_3}{(\sqrt{54.1})^3} = \frac{-26.1}{(7.4)^3} = -0.064$$

ونلاحظ أن التوزيع لـه التـواء سـالبا أي أنـه ملتـوي ناحيـة اليسـار كـما أن الالتـواء خفيفة. عندما يكون التوزيع معتدلا تكون قيمة $_3\alpha = 0$

5. حساب معامل التفرطح باستخدام العزوم

لحساب معامل التفرطح بدلالة العـزوم فأننـا نسـتخدم العـزم الرابـع حـول الوسـط الحسابي. ويرمز لمعامل التفرطح بدلالة العزم بالرمز $\alpha 4$ ويحسب من العلاقة:

$$\alpha_4 = \frac{m_4}{m_2^2}$$

مثال:

أوجد معامل التفرطح باستخدام العزوم لبيانات إنتاج الذرة بمشروع الجزيرة ؟

الحل:

لقد أوضحنا من قبل إن العزم الرابع حول الوسط الحسـابي لهـذه البيانـات يسـاوي. وعليه فأن معامل التفرطح يساوي:

$$\alpha_4 = \frac{m_4}{m_2^2} = 12902.65/(54.1)^2 = 4.41$$

وبما أن تفرطح المنحني المعتدل يساوى (3) فان المنحنى لبيانات إنتاج الذرة بمشروع الجزيرة يعتبر منحنى مدبب تدبباً كبيرا إذ انه يزيد كثيرا عن تفرطح التوزيع المعتدل.

الفصل السابع

الارتباط والانحدار

Correlation and Regression

مقدمة:

تنحصر الدراسة في مقاييس النزعة المركزية أو مقاييس التشتت في موضوع ظاهرة واحدة، أي أن البيانات أو المشاهدات أو القيم من نوع واحد تم أخذها وتسجيلها عن مجموعة من العناصر أو الافراد مثل البيانات عن أعمار الأشخاص، أو علامات طلاب، أو الدخول لمجموعة من العمال،...، وفي مثل هذه الحالات تكون الدراسة عن متغير واحد.

أما في حالة دراسة قياسين عن كل عنصر من العناصر قيد الدراسة مثل تسجيل دخول العمال وأعمارهم لندرس العلاقة بين هذين المتغيرين وهل هناك ارتباط بينهما، كأن يزيد أحدهما مع زيادة الآخر أو ينقص أحدهما إذا زاد الآخر.. إن معرفة وجود علاقة بين المتغيرين وقياس قوة تلك العلاقة هي موضوع الارتباط، وإذا كانت العلاقة خطية فإن المقياس الذي نقيس به قوة العلاقة الخطية هو معامل الارتباط الخطي.

نظرية الارتباط غالباً تظهر قوة العلاقة بين المتغيرين أو الظاهرتين (x) و (y) أما دراسة هذه العلاقة من خلال التمثيل البياني بأفضل علاقة اقتران ممكنة بالشكل (x) y=f فتسمى بدراسة الانحدار ويسمى المستقيم أو المنحنى الذي يمثل هذه الدالة بمنحنى الانحدار.

الارتباط والانحدار يعتبران من أكثر الأساليب الإحصائية فائدة واستخداماً في العلاقة بين المتغيرات في مختلف المجالات عموماً والاقتصادية والإدارية بشكل خاص. ويسمى التحليل الخطي Kinear Analysis، ويستخدم هذا التحليل من طرف الاقتصاديين والاداريين وآخرين في تحديد طبيعة وقوة العلاقة بين متغيرين أو أكثر.

وتبرز الحاجة إلى التحليل في المجالات المختلفة، إلا أن الحاجة تكون أكث مطلباً في المجالات الاقتصادية والادارية مثل دراسة العلاقة بين الايرادات والنفقات، والربح وراس المال والعمل ...

وبناء على ما تقدم فإن الارتباط والانحدار يعتبران من الاساليب الإحصائية المهمة والواسعة الانتشار والاستخدام لعرض وإيجاد العلاقة بين متغيرين إذا كانت وحدات القياس للمتغيرات من النوع المستمر. فمثلاً لدراسة حجم الانتاج في مصنع ما يمكن استخدام الارتباط والانحدار لتقدير كمية الانتاج من خلال تحديد عدد العمال أو مقدار المبلغ المخصص للأجور. وسوف نبدأ بمعامل الارتباط لتحديد العلاقة بين المتغيرين.

معامل الارتباط في دراسة العلاقة بين متغيرين، (y,x) إحداهما مستقل والآخر تابع، لمعرفة ما إذا كان تغير أحدهما مرتبط بتغير الآخر. لذلك فإن تحليلاً لارتباط يعني قياس قوة (واتجاه) العلاقة بين متغيرين أو أكثر دون التعرض لدراسة العلاقة السببية بينهما. وتبدأ دراسة الارتباط بافتراض وجود العلاقة منطقياً فمثلاً يقضي المنطق الاقتصادي بوجود علاقة بين حجم الطلب على سلعة ما وسعرها، أو حجم الانتاج والمبيعات أو الإعلان والمبيعات. ويمكن تقسيم الارتباط إلى الأنواع التالية:

1) من حيث قوة الارتباط:

* ارتباط تام كالعلاقة بين المربع وطول ضلعه ويمكن أن يظهر هذا النوع من الارتباط في العلوم الطبيعية ولكنه نادر الوجود في العلوم الاقتصادية والادارية.

* ارتباط غير تام حيث يمكن التقدير منطقياً بوجود علاقة بين المتغيرين ولكن من الصعب تفسير التغير في أحد المتغيرين كلياً بالتغير في المتغير الثاني، فمثلاً ليس حجم الدخل هو المتغير الوحيد المرتبط بحجم الطلب على سلعة ما فهناك أسعار السلع البديلة والأذواق والميول.

2) من حيث عدد المتغيرات:

* ارتباط بسيط ويدرس العلاقة بين متغيرين فقط كالارتباط بين الاعلان والمبيعات.

* ارتباط متعدد ويدرس العلاقة بين أكثر من متغيرين كارتباط بين التأمين الصحي والدرجة وعدد أفراد الأسرة.

3) من حيث الشكل:

* ارتباط خطـي/ وتمثـل العلاقـة بـين متغـيرين أو أكـثر في هـذه الحالـة يأخـذ خـط مستقيم ويعني العلاقة بـأن التغـير في أحـد المتغـيرين يكـون ثابتـاً إذا زاد المتغـير الآخـر بمقدار ثابت مهما كانت نقطة البداية كالعلاقة بين رأس المال والربح.

* ارتباط غير خطي: وتمثل العلاقة بين متغيرين أو أكثر بنمـوذج غـير خطـي، وتعنـي العلاقة بأن التغير في أحد المتغيرين إذا زاد المتغير الآخر بمقدار معين يكون غير ثابت.

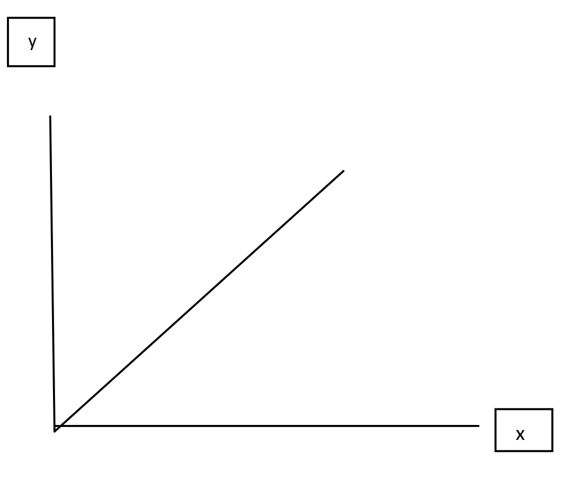
ويمكن قياس الارتباط بين متغيرين إمـا مـن خـلال رسـم شـكل الانتشـار أو القيـاس الكمي للارتباط.

قياس الارتباط:

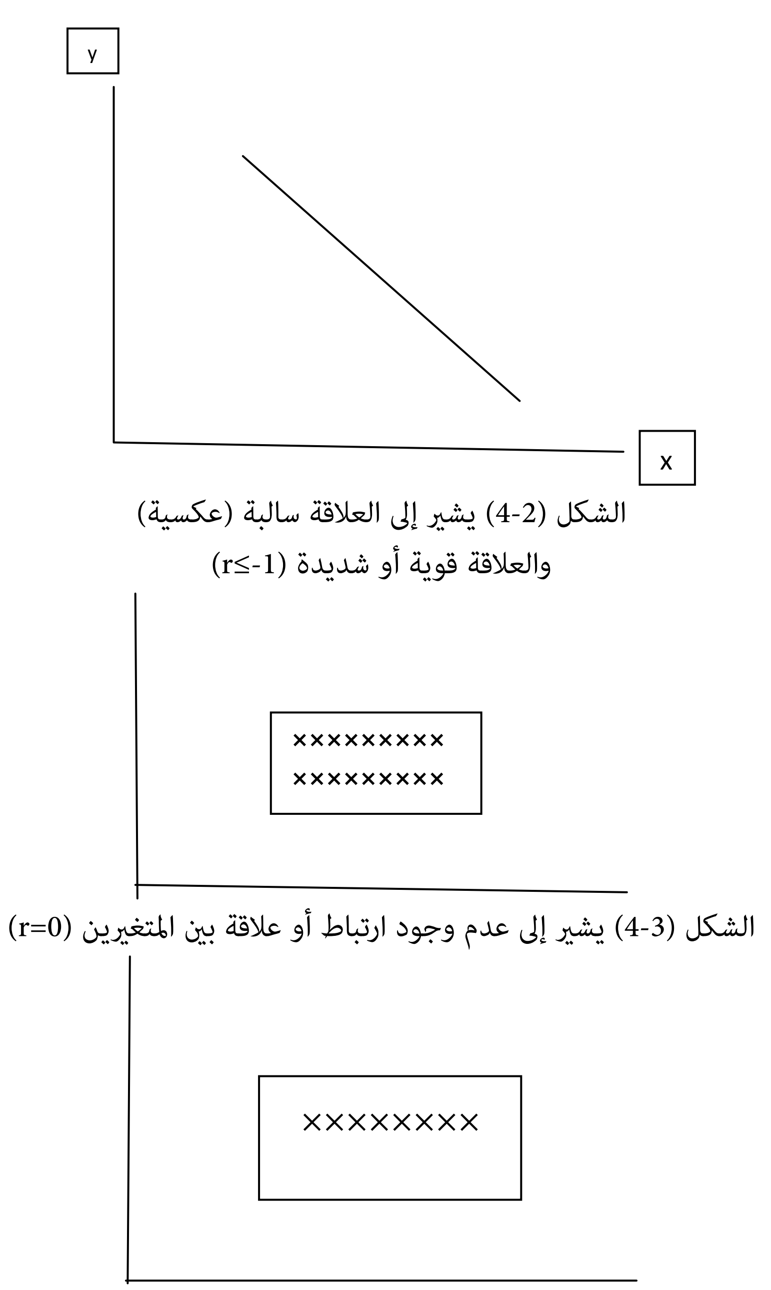
أولاً: رسم شكل الانتشار Scatter Diagram:

يعتبر من الطرق المرئية لعرض البيانات للمتغيرين وعنـد رسـم شـكل الانتشـار فقـد نحصل على أحد الاشكال التالية:

الشكل (1-4) يشير الي إن العلاقة خطية والارتباط



إيجابي (طردي) قوي أو شديد ($r \leq 1$)



الشكل (2-4) يشير إلى العلاقة سالبة (عكسية)
والعلاقة قوية أو شديدة (r≤-1)

الشكل (3-4) يشير إلى عدم وجود ارتباط أو علاقة بين المتغيرين (r=0)

(الشكل (4-4) يشير إلى أنه لا يوجد علاقة ارتباط غير خطي

وبالنظر إلى لوحة الانتار نلاحظ ما إذا كان هناك علاقة أو ارتباط بين المتغيرين (y,x) (حيث أن1≥r≥1-)، فإذا كانت العلاقة قوية وموجبة فإن

قيمة (r) تقترب من (1)، وإذا كانت قوية وسالبة فإن قيمة (r) تقـترب مـن (-1)، وكلـما اقتربت قيمة (r) من (0) (الصفر) فيعني ذلك أن العلاقة ضعيفة. ويمكن أيضـاً ملاحظـة التالي:

- إذا نقصت (y) بازدياد (x) تكون العلاقة أو الارتباط سالب.
- إذا زادت (y) بازدياد (x) تكون العلاقة أو الارتباط موجب.
- عندما لا تتأثر (y) في (x) فإذا لا يوجد علاقة.

مثال:

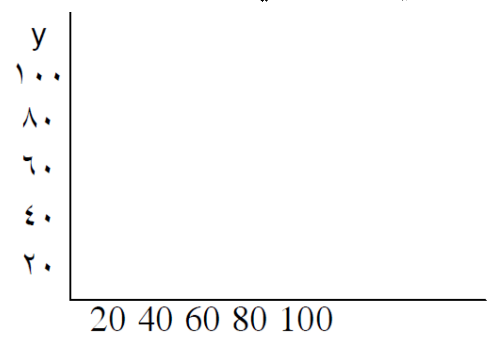
الجدول التالي يبين علامات ستة طلاب في مبحثين هما (y,x)

علامة مادة اللغة العربية (x)	علامة مادة اللغة الانجليزية (y)
69	75
85	82
75	65
90	90
80	76
50	60

والمطلوب: رسم لوحة الانتشار وتحديد نوع العلاقة بين المنهجين.

الحل:

نمثل بيانات (x,y) بياناً كما في الشكل التالي:



نلاحظ من لوحة الانتشار أن العلاقـة بـين المبحثـين (y,x) هـي علاقـة خطيـة وهـي علاقة موجبة (طردية).

ثانياً: القياس الكمي للارتباط:

إن التعرف على طبيعة وقوة الارتباط بين المتغيرين بشكل وصفي مـن خـلال لوحـة الانتشار غير كاف لهذا لابد من التعبير عن هذه العلاقة بشكل رقمي ويوجد عدة طـرق لقياس كمياً ومنها معامل ارتباط بيرسون.

معامل ارتباط بيرسون الخطي:

Pearson Linear Correlation Coefficient

معامل ارتباط برسون يقيس قوة العلاقة الخطية بين المتغيرين، ويمكن قياس العلاقة بين المتغيرين (yi,xi) إلى (n) من أزواج المشاهدات حسب قانون بيرسون التالي:

$$r = \frac{\sum(xi-x)(yi-\bar{y})}{\sqrt{\sqrt{\sum(xi-\bar{x})^2 \sum(yi-\bar{y})^2}}}$$

أوجد الصيغة التالية:

$$r = \frac{\sum xiyi - n.\bar{x}.\bar{y}}{\sqrt{\sum xi^2 - n.x^2}\sqrt{\sum yi^2 - ny^2}}$$

حيث أن:

$\Sigma xiyi$: هو حاصل ضرب القيم المتقابلة من x و y.

x^2: تربيع قيم x.

y^2: تربيع قيم y.

x: الوسط الحسابي لقيم x.

y: الوسط الحسابي لقيم y.

n: عدد البيانات أو القيم.

مثال:

أوجد معامل ارتباط بيرسون بين قيم المتغيرات التالية:

x	2	4	6	8	10	6
y	1	2	3	4	5	3

الحل:

نكون جدول جديد ونجد الوسط الحسابي لقيم (x) وقيم (y) ومن ثم نجد الفروق بين القيم وأوساطها الحسابية حسب التالي:

x	y	$xi-\bar{x}$	$yi-\bar{y}$	$(xi-\bar{x})(yi-y)$	$(\bar{x}i-x)^2$	$(yi-\bar{y})^2$
2	1	-4	-2	8	16	4
4	2	-2	-1	2	4	1
6	3	0	0	0	0	0
8	4	2	2	2	4	1
10	5	4	2	8	16	4
6	3	0	0	0	0	0
36	18			20	40	10

نجد الوسط الحسابي لقيم (x) ونجد الوسط الحسابي لقيم (yi).

$$\bar{x} = \frac{36}{6} = 6$$

$$\bar{y} = \frac{18}{6} = 3$$

وبالتعويض للصيغة التالية:

$$r = \frac{\sum(xi-x)(yi-\bar{y})}{\sqrt{\sqrt{\sum(\bar{x}i-x)^2\sum(yi-\bar{y})^2}}}$$

$$r = \frac{20}{\sqrt{40\times10}}$$

$$r = \frac{20}{\sqrt{400}}$$

$$r = \frac{20}{20}$$

إذاً $r = 1$

وهذا يعني وجود علاقة طردية وموجبة وتامة بين المتغيرين (كلما زاد x زاد y).

إن حساب قيمة (r) من المعادلة السابقة يحتاج إلى وقت وبشكل خاص إذا كانت الاوساط الحسابية تحتوي كسوراً، لذلك من الأفضل اللجوء إلى الصيغة الأخرى وهي:

$$r = \frac{\sum xiyi - n.\bar{x}.\acute{y}}{\sqrt{\sum x\bar{i}^2 - n.x^2}\sqrt{\sum yi^2 - n\bar{y}^2}}$$

والآن نحل المثال السابق لكن نكون جدول جديد كما يلي:

x	y	x.y	x^2	y^2
2	1	2	4	1
4	2	8	16	4
6	3	18	36	9
8	4	32	64	16
10	5	50	100	25
3	3	18	36	9
36	18	128	256	64

$$\bar{x} = \frac{36}{6} = 6$$

$$\bar{y} = \frac{18}{6} = 3$$

ونطبق الصيغة للمعادلة:

$$r = \frac{\sum xiyi - n.\bar{x}.\acute{y}}{\sqrt{\sum x\bar{i}^2 - n.x^2}\sqrt{\sum yi^2 - n\bar{y}^2}}$$

$$r = \frac{128 - 6(6)(3)}{\sqrt{256 - 6(6^2)}.\sqrt{64 - 6(3^2)}}$$

$$r = \frac{128 - 108}{\sqrt{256 - 216}.\sqrt{64 - 54}}$$

$$r = \frac{20}{\sqrt{40}\sqrt{10}}$$

$$r = \frac{20}{\sqrt{400}}$$

$$r = \frac{20}{20} = 1$$

وباستخدام الصيغة الثانية حصلنا على نفس الناتج وهو (r = 1) أي يعني العلاقة هي ارتباط طردي تام.

مثال (2):

احسب معامل ارتباط بيرسون بين المتغيرين (y , x) حيث قيم كل منهما كما في الجدول التالي:

x	8	1	6	4	7	7	2
Y	6	4	3	4	5	4	2

الحل:

نكون جدول جديد ونجد الاوساط الحسابية لكل من قيم (x) وقيم (y):

x	y	x.y	x^2	y^2
8	6	48	64	36
1	4	4	1	16
6	3	18	3	9
4	4	16	16	14
7	5	35	49	25
7	4	28	49	16
2	2	4	4	4
35	28	153	219	122

$$\bar{x} = \frac{35}{7} = 5$$

$$\bar{y} = \frac{28}{7} = 4$$

ونطبق h لمعادلة:

$$r = \frac{\sum xiyi - n.\bar{x}.\bar{y}}{\sqrt{\sum x\bar{i}^2 - n.x^2}\sqrt{\sum yi^2 - n\bar{y}^2}}$$

$$r = \frac{153-7(5)(4)}{\sqrt{219-7(5^2)}.\sqrt{122-7(4^2)}}$$

$$r = \frac{153-140}{\sqrt{219-175}.\sqrt{122-112}}$$

$$r = \frac{13}{\sqrt{144}\sqrt{10}}$$

$$r = \frac{13}{\sqrt{440}}$$

$$r = \frac{13}{20.98}$$

$$r = 0.62$$

وهذا يعني وجود علاقة خطية بين المتغـيرين (y, x) وهـي علاقـة موجبـة وطرديـة لكن متوسطة القوة.

وحاول عزيزي الدارس حل نفس المثال بالصيغة الأولى لترى إذا كان الجواب سـيكون نفس الشيء.

خصائص معامل ارتباط بيرسون:

1. تتراوح قيمته بين 1- , 1+.
2. يلخص اتجاه وقوة العلاقة بين المتغيرين في رقم واحد.
3. يفترض أن العلاقة بين المتغيرين (y,x) هي علاقة خطية.
4. تتأثر قيمة معامل ارتباط بيرسون بالقيم المتطرفه.
5. عندما تكون قيمة (r) تساوي (1) فهذا يعني أن الارتباط طردي وتام أمـا إذا كانـت قيمـة (r) تسـاوي (1-) فهـذا يعنـي ان العلاقـة عكسـية وتامـة وأمـا إذا كانـت (r) تساوي صفراً فهذا يعني بعدم جود علاقة بين المتغيرين.

معامل ارتباط سبيرمان للرتب:

pearman Rank Correlation Coefficient

عندما يكون من الصعب قياس المغيرات رقمياً فيفضل دراسة العلاقة بـين رتبها. كما يستخدم أيضاً في البيانـات الرقميـة لتسـهيل العمليـات الحسـابية. وعندما تكون البيانات وصفية نلجأ لتحويلها إلى بيانات عددية قابلـة للحـل.

فإذا كان لدينا مجموعة من الطلاب مسجلين في مادتين وهما مبادئ الاحصاء ومبادئ الاقتصاد الجزئي فإننا نهتم بمعرفة ما إذا كان أحسن الطلبة في مبادئ الإحصاء هم كذلك في مادة مبادئ الاقتصاد الجزئي أو كانت لديهم التقديرات مثل ممتاز جيد جداً أو (A,B,C) وفي هذه الحالة معروف الوصف وليس القيم، وللتغلب على مشاكل الاختلاف فإننا نلجأ إلى دراسة العلاقة بين رتب العلامات وليس بين العلامات نفسها.

والمقياس الذي يستخدم في هذه الحالة هو معامل ارتباط سبيرمان للرتب (rs) وتكتب صيغته بالمعادلة التالية:

$$rs = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

حيث أن:

rs: معامل ارتباط سبيرمان للرتب.

n: عدد الأزواج y , x.

d: الفرق بين xi – yi.

مثال:

البيانات التالية تمثل تقارير ستة طلاب في مادتين مبادئ الاحصاء (x) ومبادئ الاقتصاد الجزئي (y) والمطلوب حساب معامل ارتباط سبيرمان للرتب.

مبادئ الاحصاء x	ممتاز	جيد	جيد	جيد جداً	ضعيف	مقبول
الاقتصاد الجزئي y	مقبول	ضعيف	مقبول	جيد	جيد جداً	ممتاز

الحل:

نجد أولاً تراتيب البيانات المعطاة سواء كانت وصفية أو رقمية كل من المتغيرين (y , x) ثم نجد (d) والتي تعني الفرق بين رتب قيم (y , x)، ونربع (d) ومن ثم نطبق المعادلة بعد تكوين جدول جديد كمايلي:

d^3	d	رتب y	رتب x	y	x
12.25	-3.5	4.5	1	مقبول	ممتاز
6.25	2.5-	6	3.5	ضعيف	جيد

1	ذ-1	4.5	3.5	مقبول	جيد
1	1	3	2	جيد	جيد جدا
16	4	2	6	جيد جداً	ضعيف
16	4	1	5	ممتاز	مقبول
532.6	المجموع				

لايجاد رتب x أو y نتبع الخطوات التالية:

لايجاد رتب x، نرتب الوصف تصاعدياً أو تنازلياً.

x ممتاز   جيد جداً   جيد   جيد   مقبول   ضعيف

الرتب 1   2   3   4   5   6

$$\frac{4+3}{2} = 3.5$$

لايجاد رتب y، نرتب الوصف تصاعدياً أو تنازلياً

y ممتاز   جيد جداً   جيد   جيد   مقبول   ضعيف

الرتب 1   2   3   4   5   6

$$\frac{5+4}{2} = \frac{9}{2} = 4.5$$

وبعد أن تم إيجاد تربيع (d) نطبق القانون:

$$rs = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$rs = \frac{6\times 52.5}{6(36-1)}$$

$$rs = 1 - \frac{52.5}{35}$$

$$rs = 1-1.5$$

$$rs = -0.5$$

وهذا يعني وجود ارتباط عكسي متوسط القوة.

مثال(2):

جد معمل ارتباط سبيرمان للرتب بين قيم المتغيرين (y , x) كما في الجدول التالي:

x	65	90	85	90	70	65
y	85	60	75	85	55	85

الحل:

نجد رتب قيم (x) ورتب قيم (y) كمايلي:

x	90	90	85	70	65	65
الرتب	1	2	3	4	5	6

$\frac{1+2}{2} = 1.5$ $\frac{5+6}{2} = 55$

y	85	85	85	75	60	55
الرتب	1	2	3	4	5	6

$\frac{1+2+3}{3} = \frac{6}{2} = 2$

نكون جدول جديد كما يلي:

x	y	رتب x	رتب y	d	d^2
65	85	5.5	2	3.5	12.25
90	60	1.5	5	-3.5	12.25
85	75	3	4	-1	1
90	85	1.5	2	0.5	0.25
70	55	4	6	3	4
65	85	5.5	2	3.5	12.25
				المجموع	42

نطبق صيغة المعادلة التالية:

$$rs = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$rs = \frac{6\times 42}{6(36-1)}$$

$$rs = 1 - \frac{42}{35}$$

$$rs = 1-1.٢$$

$$rs = - 0.٢$$

وهذا يعني وجود ارتباط عكسي (سالب) ضعيف.

خصائص معامل ارتباط سبيرمان للرتب:

1. يستخدم في قياس العلاقة بين المتغيرات النوعية (ممتاز، جيد، ضعيف...)
2. سهل في الحساب والفهم والتطبيق.
3. يمكن استخدامه في حالة المتغيرات الكمية إذا أمكن تحويلها إلى رتب.
4. تتراوح قيمة 1+، 1-.

معامل التوافق Contingency Coefficient:

يستخدم معامل التوافق لدراسة العلاقة بين متغيرين نوعيين (y , x) وإحداهما أو كلاهما ينقسم إلى أكثر من حالتين وقد يكون إحدى أو كلا المتغيرين وصفياً وأصغر قيمة للمعامل التوافق هي الصفر وأكبر قيمة تعتمد على عدد الصفوف ولكنها لا تتجاوز (1) وصيغة احتساب معامل التوافق هي:

$$rc = \sqrt{\frac{x^2}{x^2+n}}$$

مثال:

أوجد معامل الارتباط التوافقي (rc) بين متغيري المهنة والانتاج كما في الجدول التالي:

المجموع	C	d	المهنة a	المهنة / الحالة
130	20	80	30	زيادة الانتاج

70	30	15	25	ثبات الانتاج
200	50	95	55	المجموع

نجد (x^2) باتباع التالي:

وبتطبيق صيغة المعادلة:

$$x^2 = 200\left[\frac{(30)^2}{55\times130}+\frac{(80)^2}{95\times130}+\frac{(20)^2}{50\times130}+\frac{(25)^2}{55\times70}+\frac{(15)^2}{95\times70}+\frac{(30)^2}{50\times70}\right]-200$$

$$x^2 = 200[1.59] - 200$$

$$x^2 = 231.76 - 200 = 31.76$$

$$rc = \sqrt{\frac{x^2}{x^2+n}}$$

$$rc = \sqrt{\frac{3176}{31.76+200}}$$

$$rc = \sqrt{\frac{31.76}{231.76}}$$

$$rc = \sqrt{0.137}$$

$$rc = 0.37$$

مما يدل على وجود علاقة متوسطة القوة بين المهنة وزيادة الانتاج.

معامل الاقتران Coefficient of Association:

يقيس الارتباط بين المتغيرات النوعية ويستخدم في حالة المتغيرات النوعية التي يقسم كل منها إلى وجهين فقط (2×2). إذا كان لدينا المتغيران (y,x) والبيانات المبينة في الجدول التالي:

y \ x	x_1	x_2
y_1	a_1	a_2
y_2	a_3	a_4

فيكون معامل الاقتران (ra):

$$ra = \frac{a_1.a_4 - a_2.a_3}{a_1.a_4 + a_2.a_3}$$

مثال:

جـد معامـل الاقـتران (ra) بـين ظـاهرتي التـدخين والمسـتوى التعليمـي لعينـة مـن الاشخاص عددها (120) وكما في الجدول:

	غير متعلم	متعلم
مدخن	35	30
غير مدخن	15	40

الحل:

$$ra = \frac{35.40 - 15.30}{35.40 + 15.30}$$

$$ra = \frac{1400 - 450}{1400 - 450}$$

$$ra = \frac{950}{1850} = 0.513$$

ويعني وجود علاقة متوسطة القوة.

الانحدار Regression:

يهتم الباحث في مجال الاقتصاد والادارة أو العلـوم الأخـرى إلى صـياغة نمـوذج يمثـل العلاقة بين المتغيرات باستخدام الطرق الإحصائية وتحتاج صياغة النموذج تحديد المتغـير المستقل (Independent variable) ويرمز له بالرمز (x) والمتغير التـابع (Dependent Variable) ويرمز له بالرمز (y). وقد يكون النموذج خطياً أو غير خطي، كما نأخذ نماذج الانحـدار أشـكالاً مختلفـة ويـتم تقـديرها واختبـار الفرضـيات المتعلقـة بهـا وتحليلهـا واستخدامها في التنبؤ بقيم المتغير التـابع عنـد مسـتويات محـددة للمتغـيرات المسـتقلة. ويعتبر التنبؤ من أهم التطبيقات الإحصائية في الاقتصاد والإدارة وغيرها.

فالانحدار هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بـين المتغـيرين (y,x) تستعمل للتنبؤ عن قيم سابقة وقيم مسـتقبلية لكـن مـن (y,x) حسـب المعلـوم مـنهما وتكون هذه المعادلة خطية، وقد تكون بدرجة ثانية أو ثالثة ولكن الاهتمام سيكون على الخطية منها فقط.

فعلى سبيل المثال اريد التنبؤ بالطلب على سلطة ما بناءً على معرفتنا بسعرها. إن التنبؤ في الطلب على السلعة لا يعتمد فقط على سعرها، وإنما على عوامل أخرى أيضاً مثل الأذواق والميول وأسعار السلع البديلة والجودة.. ولكن لتبسيط المثال نفترض وجود متغيرين وهما سعر السلعة وهو (x) ومن معرفتنا لقيم (x) نريد أن نتنبأ لقيم (y) (الطلب على السلعة). وإذا عرفنا قيمة معينة للمتغير (x) نستطيع تعويضها في معادلة الانحدار لنجد تنبؤ قيمة (y) للمتغير (y) المقابلة للقيمة المعطاة.

معادلة انحدار متغير على متغير آخر:

إذا كان لدينا عينة من الأزواج المرتبة كما يلي:

$$(x_1.y_1),(x_2,y_2)\ ,\ (x_3,y_3)\ldots(x_n,y_n)$$

حيث $(x_1,x_2,x_3,\ldots,x_n)$ قيم المتغير (x) وحيث أن $(y_1,y_2.y_3,\ldots,y_n)$ القيم المقابلة للمتغير المستقل (x) وتسمى قيم المتغير (y) (التابع). وإذا رصدنا هذه النقاط على المستوى (xy) نحصل على لوحة الانتشار ونتمكن من الحكم فيما إذا كان من الممكن تطبيق خط مستقيم على شكل الانتشار أم لا.

فإذا فرضنا أن هناك علاقة خطية بين المتغيرين (y,x) فيمكن التعبير عن هذه العلاقة بالصيغة التالية:

$$Y = A + BX + e$$

حيث (e) متغيرات عشوائية وربما ليس لها قيمة أو تأثير.

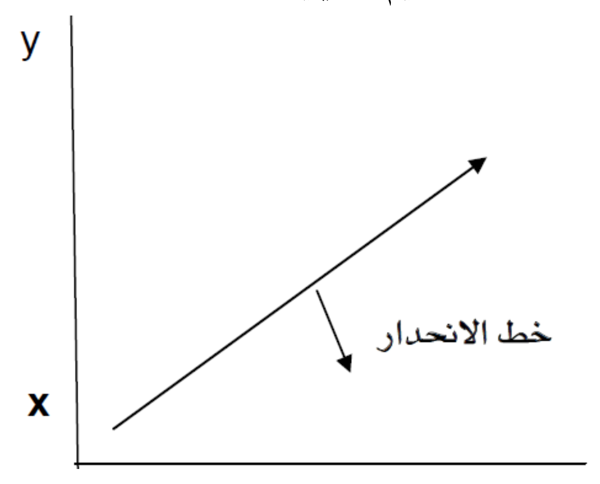
والمطلوب: هو تقدير (A,B) حتى نستطيع تقدير قيمة (y_1) المقابلة للمتغير (x_1).

لهذا نفرض ان (a) هو تقدير ((b ،A)) هو تقدير (B) ولذلك تكون الصيغة الجديدة التالية:

$$\bar{y} = a + bX$$

وهو خط انحدار (y) على (x) وحصلنا عليه باستعمال (b,a) ويكون ($\bar{y}$) هي القيمة التقديرية المتنبأة بها ويكون ($e=y-\bar{y}$) هو الخطأ في التقدير.

الشكل التالي يمثل لوحة انتشار لقيم المتغيرين (y , x):



من المنطق أن يكون الخط المستقيم (خط الانحدار) لوحة الانتشار يمثل النقاط أحسن تمثيل.

إن إحدى الطرق التي تستعمل لايجاد ذلك الخط المستقيم هي طريقة المربعات الصغرى.

طريقة المربعات الصغرى Least Squares Method:

هي طريقة تطبيق خط مستقيم على مجموعة من النقاط للازواج (xiyi) بحيث يكون مجموع مربعات الأخطاء أصغر ما يمكن وحتى يتم حل معادلة خط الانحدار:

$$\bar{y} = a+bX$$

لابد من معرفة قيمة (b) حتى تمكن من معرفة قيمة (a) وبالتالي المعادلة وعليه:

$$b = \frac{\sum x.y - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$a= \bar{y} - b\bar{X}$$

علماً بأن (X) هو الوسط الحسابي لقيم $(x_1,x_2,x_3,\dots,x_n)$ و $(\bar{y})$ هو الوسط الحسابي لقيم $(y_1,y_2.y_3,\dots,y_n)$ و (n) عدد القيم أو المشاهدات وبتعويض قيم (a,b) في المعادلـة $(\bar{y}=a+bX)$ نحصل على معادلة انحدار (y) على (x).

مثال:

جد معادلة انحدار (y) على (x) للبيانات التالية وبماذا تقدر قيمة (y) إذا كانت (x) المقابلة لها تساوي (5).

x	0	4	2	3	1
y	1	6	2	5	1

الحل:

نكون جدول جديد وعلى النحو التالي:

x	y	x.y	x^2
0	1	0	0
4	6	24	16
2		4	4
3	5	15	9
1	1	1	1
10	15	44	30

$$x = \frac{10}{5} = 2$$

$$\bar{y} = \frac{15}{5} = 3$$

نجد قيمة (b) من خلال المعادلة:

$$b = \frac{\sum x.y - \overline{nxy}}{\sum x^2 - n\bar{x}^2}$$

$$b = \frac{44-5(2)(3)}{30-5(2)^2}$$

$$b = \frac{44-30}{30-20}$$

$$b = \frac{14}{10} = 1.4$$

وجد قيمة (a) من الصيغة:

$$a = \bar{y} - b\breve{X}$$

$$a = 3-1.4(2)$$

$$a = 3-2.8 = 0.2$$

وبذلك تكون معادلة الانحدار هي:

$$\bar{y} = 0.2 + 1.4X$$

وإذا كانت (X-5) فإن قيمة (y) المتنبأ بها هي:

$$\bar{y} = 0.2 + 1.4(5)$$

$$\bar{y} = 0.2 + 7 = 7.2$$

وهذا يعني عندما تكون (X = 5) فإن (y = 7.2).

مثال (2):

الجدول التالي يمثل العلاقة بين متغيرين (y , x)

x	4	10	9	12	8	5
y	2	6	8	11	5	4

والمطلوب:

1. أوجد معادلة الانحدار (y) على (x).
2. بماذا تقدر قيمة (y) عندما تكون قيمة (x) تساوي (9).
3. ما الخطأ في تقدير (y) عندما تكون (x) تساوي (9).

الحل:

نكون جدول جديد كمايلي:

x	Y	x.y	x^2
4	2	8	16
10	6	60	100
9	8	72	81

12	11	132	144
8	5	40	64
5	4	20	25
48	36	332	430

$$\bar{x} = \frac{48}{6} = 8$$

$$\bar{y} = \frac{36}{6} = 6$$

بعد أن وجدنا الوسط الحسابي لقيم (y,x) نجد قيمة (b) من الصيغة التالية:

$$b = \frac{\sum x.y - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$b = \frac{332-6(8)(6)}{430-6(8)^2}$$

$$b = \frac{332-288}{430-384}$$

$$b = \frac{44}{46} = 0.96$$

ونجد قيمة (a) من الصيغة التالية:

$$a = \bar{y} - b\bar{X}$$

$$a = 6-0.96(8)$$

$$a = 6-7.68 = -1.68$$

1) إذا المعادلة هي:

$$\bar{y} = -1.68 + 0.96(X)$$

2) القيمة التقديرية المتنبأ بها للمتغير (y) عندما تكون (x) تساوي (9) هي:

$$\bar{y} = -1.68 + 0.96(9)$$

$$\bar{y} = -0.168 + 8.64 = 6.96$$

3) الخطأ في تقدير (y) عندما تكون (x) تساوي (9) هي:

$$e = y - \bar{y}$$

قيمة ($\bar{y}$) المتنبأ بها تساوي (9.96) وقيمة (y) هي الحقيقية في الجـدول وننظـر إلى قيم المتغير (x) ونجد أن (x=9) في الجدول يقابلها (y) يساوي (8).

وعليه الخطأ في التنبؤ:

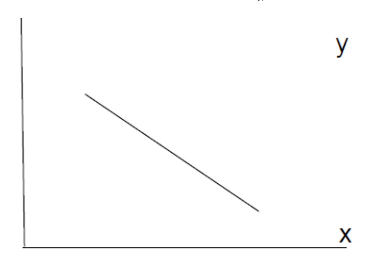
$$e = y - \bar{y}$$

$$e = 8 - 6.96$$

$$e = 1.04$$

تمارين الفصل السابع

1. وضح المقصود بكل من الارتباط والانحدار.
2. وضح الارتباط الخطي البسيط من حيث الشكل.
3. إلى ماذا يشير الشكل التالي:



4. أوجد معامل ارتباط بيرسون بين قيم المتغيرات التالية: ك

x	4	6	5	9
y	3	7	3	5

5. إذا كان مجموع x,y(420) والوسط الحسابي x(8) والوسيط الحسابي $\bar{y}$(6) ومجموع x2(510) ومجموع القيم هو (9). فما هو معامل ارتباط بيرسون بين المتغيرين.
6. البيانات التالية تمثل تقارير خمسة طلاب في مادتين الرياضيات والادارة والمطلوب حساب معامل ارتباط سبيرمان للرتب.

الرياضيات	C	A	D	C	F
الاداره	B	C	D	B	C

7. جد معامل ارتباط سبيرمان للرتب بين قيم المتغيرين (y,x)

x	165	170	190	185	190	165
y	185	155	185	175	160	185

8. أذكر خصائص معامل ارتباط سبيرمان.

9. أوجد معامل الارتباط التوافقي بيم متغيري زيادة الاجور والانتاج كما في الجدول التالي:

الانتاج / زيادة الأجور	المصنع أ	المصنع ب	المصنع ج
زيادة الاجور	40	50	15
ثبات الاجور	15	20	10

10. جد معامل الاقتران بين ظاهرتي امتلاك رخصة قيادة والمستوى التعليمي لمجموعة من الاشخاص.

أمي	متعلم	المستوى التعليمي / الحالة
15	28	يمتلك رخصة
19	10	لا يمتلك رخصة

11. الجدول التالي يمثل العلاقة بين المتغيرين y , x

8	20	18	24	16	10	x
4	12	16	22	10	8	y

المطلوب:

أ-أوجد معامل خط الانحدار.

ب-بماذا تقدر قيمة y عندما تكون x=18.

ج-ما الخطأ في تقدير y عندما تكون x =18.

الفصل الثامن

السلاسل الزمنية

Time Series

مقدمة:

سوف نتطرق في هذا الفصل إلى دراسة المشاهدات أو البيانات المأخوذة لعدد من الفترات الزمنية والتي تسمى السلاسل الزمنية وذلك عن طريق وصف هذه السلاسل ومن ثم تحليلها.

وسوف يتم التركيز على الاتجاه العام وكيفية تحديده سواء من خلال الانتشار باليد أو نصف السلسلة أو عن طريق المربعات الصغرى.

تعريف السلسلة الزمنية:

هي عبارة عن مجموعة من البيانات أو المشاهدات أو القياسات الإحصائية التي يتم جمعها عن ظاهرة ما، وعلى فترات زمنية متعددة. وغالباً مايكون هذا الترتيب فيه تساوي الفترات الزمنية مثل الأيام، الأشهر أو السنوات مثل معرفة اسعار النفط خلال فترة من الزمن أو استهلاك النفط على مدار أشهر السنة.

وغالبا ماتكون قياسات السلاسل الزمنية الأسعار أو الكميات، ولكن ما يميزها هو الترتيب للفترات الزمنية فمثلاً ان السلسلة تمثل الاسعار للفترات (n...4,3,2,1) أو الاسعار للسنوات (2000، 2001 ، 2002).

وتنقسم السلاسل الزمنية إلى:

1) الاتجاه العام أو التغيرات الاتجاهية:

وتفيدنا في عملية التخطيط طويل الأجل، وذلك في حالة الظواهر التي تؤثر فيها عوامل ثابتة لها صفة الاستمرارية مثل النمو السكاني.

2) التغيرات الموسمية (الفصلية):

ويقصد بها تلك التي تظهر في موسم معين دون غيره، مثلاً يزيد أو من المحتمل أن يزيد دخل الدولة في فصل الصيف من العملات الاجنبية. أو يزداد الاستهلاك في المشتقات النفطية في فصل الشتاء.

3) التقلبات أو التغيرات الدورية:

وهي تلك التي تظهر في بداية كل دورة معينة ثم تختفي لتعود بعد ذلك من جديد في بداية الدورة التالية. مثل زيادة الانتاج في بداية الاسبوع أو عدد صفحاتها يكون قليل في يوم العطلة. أو زبادة حجم الودائع لدى البنوك في اليوم الاخير من الاسبوع.

4) التغيرات العرضية:

وهي ليست مرهونة بمكان أو زمان وإنما مرهونة بظروف طارئة. مثل شراء وتخزين المواد التموينية عندما يسمع المواطنون بإمكانية حدوث ظروف سيئة في يوما ما. (تساقط الثلوج) ـو حدوث حرب).

تخضع السلسلة الزنية لمجموعة من التأثيرات تسمى مركبات السلسلة الزمنية وأهمها مركبة الاتجاه العام والمركبة الموسمية ومركبة ادورة ومركبة الخطأ. وفي هذا الفصل سنركز على مركبة الاتجاه العام لأنها الأكثر أهمية.

مركبة الاتجاه العام:

وهي المركبة الأكثر أهمية يمكن تقديرها بإحدى الطرق التالية:

1) طريقة الرسم أو الانتشار باليد:

وتتلخص خطوات هذه الطريقة بان نقوم برصد النقاط الظاهرة مع الزمن وبعد ذلك محالة رسم خط مستقيم (الاتجاه العام) أو منحنى يمثلها أفضل تمثيل ويمر في غالبية نقاطها.

أما إذا كانت النقاط مبعثرة تماماً بحيث يصعب رسم خط يمثلها، في هذه الحالة لا يكون لهذه الظاهرة إتجاه عام.

إن دقة الرسم بهذه الطريقة تعتمد على نوعية البيانات من طرف وعل مهارة الشخص الذي سيقوم برسم خط الاتجاه العام من طرف آخر. ولكنها قد تمنحنا فكرة مبدئية عن الظاهرة المراد دراستها وقد تمكنا من التنبؤ للمستقبل.

ويمكن توضيحها من خلال المثال التالي:

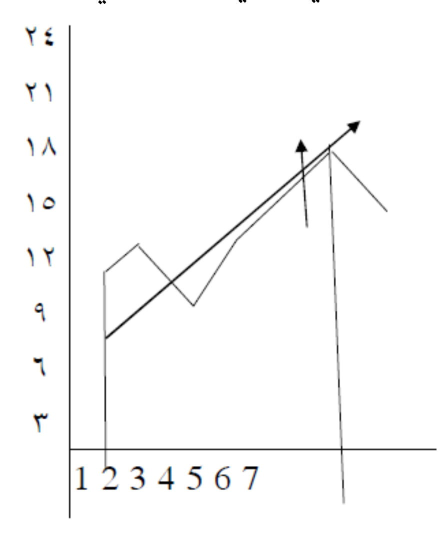
مثال (6-1):

البيانات التالية تمثل قراءة تمثل قراءة درجة الحرارة لأحد الاسابيع في فصل الربيع والمطلوب التمثيل البياني وإيجاد خط الاتجاه العام ومن ثم معالة خط الاتجاه العام.

اليوم	سبت	أحد	أثنين	ثلاثاء	اربعاء	خميس	جمعة
بالارقام	1	2	3	4	5	6	7
المشاهدات	12	15	10	18	20	22	17

الحل:

نرصد البيانات على المستوى البياني كما في الشكل التالي:



- لايجاد خط الاتجاه العام والمعادلة نأخذ نقطتان وتمثلان النقاط أحسن تمثيل ولتكن (1.12)، (6، 22).

- نجد معالة خط الاتجاه العام من العلاقة التالية:

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y - y_1}{x - x_1}$$

وبالرجوع إلى النقطتان (1.12) و (6.22) نطبق المعادلة فتكون:

$$\frac{22-12}{6-1} = \frac{y-12}{x-1}$$

$$\frac{10}{5} = \frac{y-12}{x-1}$$

$$\frac{2}{1} = \frac{y-12}{x-1}$$

$$y = 2(x-1) + 12$$

$$y = 2x + 10$$

وتعتبر هذه الطريقة غير دقيقة وتختلف من شخص إلى آخر وذلك بسبب اختيار النقطتان ولكن مع عدم الدقة إلا أنها تعطي فكرة سريعة عن طبيعة البيانات.

2) طريقة المعدل النصفي أو متوسط نصف السلسلة:

Semi – Average method

تعتبر هذه الطريقة أكثر دقة من طريقة الانتشار باليد وتتلخص خطواتها بالتالي:

1. نرقم قيم المشاهدات بأرقام متسلسلة (...,3,2,1)
2. نأخذ المتوسط الحسابي للنصف الأول من ترتيب المشاهدات فيكون هو الإحداث السيني لنقطة الأولى. كما نأخذ المتوسط الحسابي للنصف الثاني من ترتيب المشاهدات فيكون هو الإحداث السيني للنقطة الثانية.
3. إذا كان عدد المشاهدات فردياً تهمل القيمة الوسطى.
4. نأخذ المتوسط الحسابي للنصف الأول من قيم المشاهدات فيكون هو الاحداثي الصادي للنقطة الاولى. كما نأخذ المتوسط الحسابي للنصف الثاني من قيم المشاهدات فيكون هو الاحداثي الصادي للنقطة الثانية.
5. نعين النقطتين على المستوى الاحداثي، ونصل بينهما بينهما ليكون هو خط الاتجاه العام.
6. نجد معادلة خط الاتجاه العام من العلاقة التالية:

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y - y_1}{x - x_1}$$

مثال (2):

البيانات التالية تمثل الانتاج السنوي لمصنع سيارات مقدرة بـآلاف السيارات ولستة سنوات.

والمطلوب: إيجاد معادلة خط الاتجاه العام والتمثيل البياني بطريقة نصف السلسلة الزمنية.

السنة	1995	1996	1997	1998	1999	2000
المشاهدات عدد السيارات (آلاف)	12	13	11	15	18	21

الحل: نكون جدول جديد كما يلي:

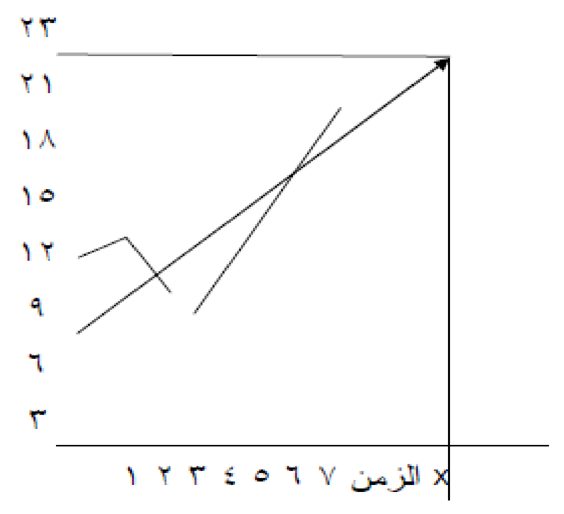
السنة	بالارقام (x)	انتاج السيارات (y)	متوسط x	متوسط y
1995	1	12	$x_1 = \frac{1+2+3}{3}$	$\bar{y}1 = \frac{12+13+11}{3}$
1996	2	13		
1997	3	14	$x1 = 2$	$\bar{y}1 = 12$
1998	4	15	$x_2 = \frac{4+5+6}{3}$	$\bar{y}_2 = \frac{12+13+11}{3}$
1999	5	18		
2000	6	21	$x_2 = 5$	$\bar{y}_2 = 18$

إذن النقطتان هما:

$$x_1 = 2 , x_2 = 5$$

$$y_1 = 12, y_2 = 18$$

ومن ثم نقوم بالتمثيل البياني كما يلي: (y) المشاهدات



وعليه تكون معادلة خط الاتجاه العام كما يلي:

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y - y_1}{x - x_1}$$

$$\frac{18 - 12}{5 - 2} = \frac{y - 12}{x - 2}$$

$$\frac{6}{3} = \frac{y - 12}{x - 2}$$

$$\frac{3}{1} = \frac{y - 12}{x - 2}$$

$$y - 12 = 2(x-2)$$

$$y - 12 = 2 \times -4$$

$$y = 2x + 8$$

وإذا أردنا توقع التنبؤ بإنتاج عام (2001) فيكون الترتيب له رقم (7) وبالتالي:

انتاج 2007 = y = 2(7) + 8 = 22

بالرسم البياني نأخذ خط مستقيم متقطع نحو الأعلى حتى يتلامس مع خـط الاتجـاه العام، ثم نأخذ خط آخر أفقي نحـو المشـاهدات فيكـون لـدينا المشـاهدة (22) وبالتـالي نتوقع بأن إنتاج عام (2007) هو (22) ألف سيارة.

3) طريقة المربعات الصغرى Least sum of squares:

تعتبر طريقة المربعات الصغرى أدق من الطريقتين السابقتين وهما الانتشار ونصف السلسلة الزمنية ومعادلتها هي:

$$T = \bar{y} = a+bx$$

وهي أن نجد معادلة خط الانحدار والتي تم شرحها سابقاً.

- السلسلة: عبارة عن مشاهدات أو بيانات عن ظاهرة خلال الزمن.

بما أن السلسلة الزمنية عبارة عن مشاهدات أو بيانات (y) مقابل الزمن (x)، فإنه يجب تعين نقطة المركز أو الاصل، أي تعين سنة محددة يكون (x=0)، بمعنى آخر تكون نقطة الأصل (سنة المركز) مقابلها طرح (1) حسب التي تلي المركز أو قبله وهكذا... ولتوضيح ذلك نأخذ المثال التالي:

مثال (3):

اذا كانت لدينا البيانات كما في الجدول التالي:

السنة	2000	2001	2002	2003	2004	2005	2006
الانتاج	20	30	32	23	34	39	32

والمطلوب:

- إيجاد خط الاتجاه العام وتمثيله بيانياً بطريقة المربعات علماً بأن سنة المركز هي (2000).
- كم تقدر إنتاج (2007).

الحل: نكون جدول جديد كما يلي:

السنة	x	الانتاج (y)	x.y	x^2
2000	0	20	0	0
2001	1	30	30	1
2002	2	32	64	4
2003	3	23	69	9
2004	4	34	136	16
2005	5	39	195	25
2006	6	32	192	36
المجموع	21	210	686	91

نجد الوسط الحسابي لكل من (y,x):

$$\bar{x} = \frac{21}{7} = 3$$

$$\bar{y} = \frac{210}{7} = 30$$

ثم نجد معالة الانحدار:

$$T = \bar{y} = a + bx$$

لذلك نجد أولاً (b)

$$b = \frac{\sum x.y - n.x.\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$b = \frac{686 - 7(3)(30)}{91 - 7(3)^2}$$

$$b = \frac{686 - 630}{91 - 63}$$

$$b = \frac{56}{28} = 2$$

$$b = 2$$

ثم نجد (a) من خلال:

$$a = \bar{y} - bx$$

$$a = 30 - 2(3)$$

$$a = 30 - 7$$

$$a = 24$$

إذن تكون معادلة الانحدار:

$$T = \bar{y} = a + bx$$

$$T = \bar{y} = 24 + 2x$$

ولرسم معادلة خط الانحدار بيانيا نفرض نقطتين كما يلي:

$$x_1 = 1 \text{ , } x_2 = 4$$

$$y_1 = 26, y_1 = 32$$

إذن النقطتان هما (4,32)، (1.26)

ومن ثم نمثل جميع النقاط بيانياً ونجد خط الاتجاه العام كما يلي:

$$b = \frac{56}{28}$$

$$b = 2$$

$$a = \bar{y} - bx$$

$$a = 30-2.(0)$$

إذن المعادلة تكون:

$$T = \bar{y} = a + bx$$

ولتغير مركز الاتجاه سحبنا المركز (k) من الوحدات فإن المعادلة الجديدة تصبح:

$$T = \bar{y} = a + b\,(x+k)$$

مع ملاحظة التالي تكون (k) موجبة إذا غير المركز إلى العام الذي يلي عام المركز الأصلي، وتكون سالبة إذا غير المركز إلى عام يسبق المركز الاصلي. لذلك في المثالين السابقين عندما كان المركز عام (2000) كانت المعادلة:

$$T = 24 + 2x$$

وعندما غيرنا المركز إلى (2003) أي أن (k=3) فأصبحت المعادلة الجديدة بعد تغير مركز الاتجاه كما يلي:

$$T = \bar{y} = 24 + 2(x+3)$$

$$T = \bar{y} = 24 + 2x + 6$$

$$T = \bar{y} = 30 + 2x$$

ولتحويل خط التجاه السنوي إلى خط اتجاه شهري، يتم ذلك بقسمة (a) على (12)، و (b) على $(12)^2$ وبذلك تكون المعادلة:

$T_1 = a + bx$ هي الوحدة بالسنة x

$T_2 - \frac{b}{144} + \frac{a}{12}$ -X هي الوحدة بالشهر x

4) طريق المعدلات (المتوسطات) المتحركة Moving Averages:

تكمن أهمية المعدلات المتحركة في أنها تعمل على الحد من خشونة السلسلة وجعلها ملساء. وتقوم هذه الطريقة على إيجاد المتوسطات الحسابية لمجموعة متتابعة ومتداخلة من قيم الظاهرة بقصد إزالة بعض التعرجات التي قد تظهر في الخاصة بها. وقد تكون كل من هذه المجموعات المتتابعة مكونة من سنتين أو ثلاثة أو أربعة... الخ

المعدلات المتحركة بطول فردي:

إذا كان طول السلسلة الزمنية بطل فردي (3) فإن المعدلات تكون كما يلي:

المتوسط الاول = $\frac{3x_1+x_2+x_3}{3}$

المتوسط الثاني = $\frac{x_2+x_3+x_4}{3}$

المتوسط الثالث = $\frac{x_3+x_4+x_3}{3}$

وهكذا حتى نحصل على عناصر السلسلة الجديدة.

مثال (5):

جد المعدلات المتحركة بطول فردي (3) للسلسلة الزمنية في الجدول التالي:

السنة	2000	2001	2002	2003	2004	2005	2006
المشاهدات	15	12	9	18	15	24	27

الحل:

نكون جدول جديد كما يلي:

السنة	المشاهدات	مجموع الشاهدات بطول 3	معدل المشاهدات بطول 3
2000	15	-	-
2001	12	36	12
2002	9	39	13
2003	18	42	14
2004	15	57	19
2005	24	66	22
2006	27	-	-

نحدد موقع المعدل الأول من العلاقة التالية:

موقع المعدل المتحرك الأول = $\frac{\text{طول المدة}+1}{2}$

$\frac{3+1}{2} = 2$

أي أن المعدل المتحرك يقابل المشاهدة الثانية (2001) في السلسلة الزمنية ثم نجد مجموع المشاهدات بطول (3) كما يلي:

$$1)\frac{15+12+9}{3} = 36$$

$$2)\frac{12+9+18}{39} = 39$$

$$3)\frac{9+18+15}{3} = 42$$

$$4)\frac{18+15+24}{3} = 57$$

$$5)\frac{15+24+27}{3} = 66$$

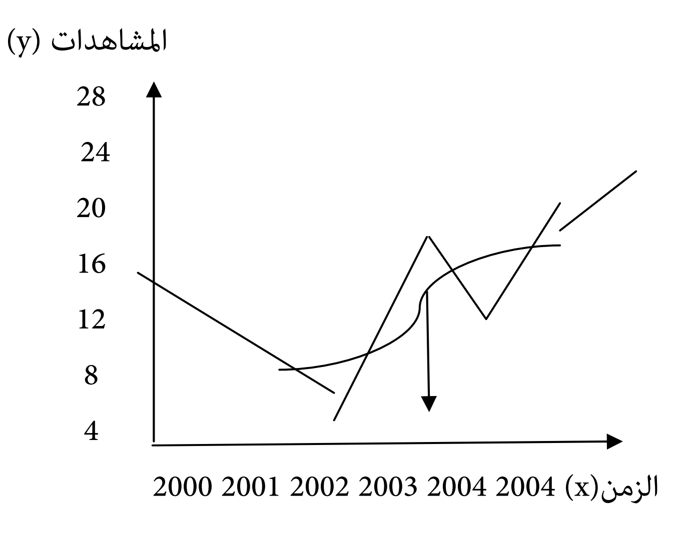
6)24.27

لا نأخذها لأنها ليس بالطول (3) المطلوب.

ثم نجد المعدل المتحرك بطول ذلك وذلك بقسمة مجموع المشاهدات بطول (3) على (3):

1. $\frac{36}{3} = 12$
2. $\frac{39}{3} = 13$
3. $\frac{42}{3} = 14$
4. $\frac{57}{3} = 19$
5. $\frac{66}{3} = 22$

أن العمود الاخير يمثل المعدلات المتحركة بطول (3) سنوات وعندما نرصد هذه المعدلات على نفس الشكل الانتشار للسلسلة الزمنية نلاحظ كيف ثم تمهدها فاصبحت ملساء نوعاً ما وتعينا هذه المعدلات اتجاه السلسلة الزمنية ويكون شكلها كما في الشكل التالي:



المعدلات المتحركة بطول زوجي:

إذا كان طول السلسلة الزمنية بطول زوجي (4) فإن المعدلات تكون كما يلي:

المتوسط الاول = $\frac{3x_1+x_2+x_3+x_4}{4}$

المتوسط الثاني = $\frac{x_2+x_3+x_4+x_5}{4}$

وهكذا حتى نحصل على عناصر السلسلة الزمنية الجديدة.

مثال (6-6):

أوجد المعدلات المتحركة بطول (4) للسلسلة الزمنية المعطاه في الجدول التالي:

تقدير المركبة الفصلية Measurment of Seasone Variation:

لحساب المركبة المفصلية للمثال السابق نجد المعدل المركزي بطـول (4) ثـم نضـيف عمود (I×S) مركبتي الفصل والخطأ ونجد المعدل المركزي كما يلي:

الاول $8.75 = \frac{9+8.5}{2}$

الثاني $8.75 = \frac{8.5+9}{2}$

الثالث $9.25 = \frac{9+9.5}{2}$

وهكذا للباقي. أما لحساب $(S\times I)$:

$$S \times I = \frac{\text{المشاهدة الاصلية}}{\text{المعدل المتحرك}} \times 100\%$$

ويكون الجدول التالي:

السنة	الفصل	المشاهدات	المعدل بطول (4)	المعدل المركزي بطول (4)	S×I
2000	I	8	-	-	-
	II	10	-	-	-
	III	6	9	8.75	6/8.75 = 68.57
	IV	12	8.5	8.75	12/8.75 = 137.14
2001	I	6	9	9.25	6/9.25 = 64.86
	II	12	9.5	9/75	12/9.75= 123.08
	III	8	10	10.25	8/10.25=78.05
	IV	14	10.5	10.25	14/10.25=136.6
2002	I	8	10	10.25	8/10.25=78.05
	II	10	10.5	10.50	10/10.5=95.24
	III	10	10.5	-	
	IV	14	-	-	

وبعد ذلك نعيد ترتيب مشاهدات مركبتي الفصل والخطأ في الجدول التالي:

الفصل السنة	الربع الأول	الربع الثاني	الربع الثالث	الربع الرابع	
2000	-	-	68.57	137.14	
2001	64.86	123.08	78.05	136.6	
2002	78.05	95.24	-	-	
المجموع	142.73	218.32	146.62	273.74	
المعدل	71.45	109.16	73.31	136.87	390.8

وبعد ذلك نجد مركبات الفصل كما يلي:

مركبة الفصل الأول:

$$\frac{71.45}{390.8} \times 400 = 73.17$$

مركبة الفصل الثاني:

$$\frac{109.16}{390.8} \times 400 = 111.73$$

مركبة الفصل الثالث:

$$\frac{73.31}{390.8} \times 400 = 75.03$$

مركبة الفصل الرابع

$$\frac{136.87}{390.8} \times 400 = 140.10$$

إن تفسير مركبة الفصل كما يلي:

إن عدد المشاهدات في الربع الأول الأشهر الثلاثة الأولى من السنة يساوي (73.17) بالمائة من معدل عدد المشاهدات في الربع الواحد.

تمارين الفصل الثامن

1. عرف السلسلة الزمنية وأذكر أقسامها.

2. البيانات التالية تمثل أعداد السياح لمدينة البتراء خلال ستة أشهر.

الشهر	1	2	3	4	5	6
المشاهدات	2400	1900	3000	3600	4000	4800

المطلوب:

أ.تمثيل هذه المشاهدات بيانياً.

ب.إيجاد خط الاتجاه العام.

ج.إيجاد معاداة خط الاتجاه العام.

3. البيانات التالية تمثل مبيعات مصنع زيوت خلال سبعة اعوام. المطلوب: إيجاد معادلة خط الاتجاه العام والتمثيل البياني لهذه المشاهدات بطريقة متوسطة السلسلة الزمنية وكم تقدر مبيعات 2007.

السنة	2000	2001	2002	2003	2004	2005	2006
المبيعات (آلاف الدنانير)	120	150	180	220	260	300	350

4. حسب معطيات السؤال السابق جد خط الاتجاه العام بطريقة المربعات الصغرى مع العلم بان سنة الاساس هي 2002، وبكم تقدر انتاج 2008؟

5. جد المعدلات المتحركة بطول (3) للسلسلة الزمنية كما في الجدول التالي:

السنة	1990	1991	1992	1993	1994	1995
المشاهدات	7	6	5	9	15	22

6.جد المعدلات المتحركة بطول (4) ومثلها بيانياً للسلسلة الزمنية كما في الجدول التالي:

السنة	2000	2001	2002	2003	2004	2005	2006
المشاهدات	4	5	3	6	3	6	4

الفصل التاسع

الارقام القياسية

Index Numbers

مقدمة:

الارقام القياسية هي أداة لوصف المتغيرات الاقتصادية وتستعم لوصف التغير في الاسعار او الكميات أو القيمة مع الزمن وبالتالي نستطيع من خلالها التعرف على الارتفاع في المستوى العام للأسعار ومعدل التضخم. والرقم القياسي لتكاليف المعيشة يعطينا نسبة التغير في تكلفة المعيشة الحياتية بين عام وآخر مقاس بتغير أسعار سلة من المواد من وقت لآخر، وعلى سبيل المثال إذا كان الرقم القياسي لتكاليف المعيشة في بلد ما وصل إلى 140% في عام (2000) مقارنة مع عام (1998) فإن هذا يعني زيادة الإنفاق عام (2000) لشراء مجموعة من السلع عن عام (1998) بما نسبته 40$ وبمعنى آخر ارتفاع تكاليف المعيشة بنسبة 40% فإذا كانت عائلة تنفق مبلغ وقدره (200) ينار عام (1998) فإنها سوف تحتاج مبلغ وقدره (280) للانفاق على نفس الحاجات عام (2000).

ويمكن الاستفادة من الأرقام القياسية لقياس التغير في البطالة والإنتاجية ومعدلات الأجور ومقارنة الأسعار فالرقم القياسي هو أداة إحصائية مصممة لتظهر التغير في قيمة الظاهرة أو مجموعة مرتبطة من الظواهر قيد الدراسة والتي لها علاقة بالنسبة لقيمتها في الزمن والمكان الجغرافي أو أية خاصية أخرى. ويعرف الرقم القياسي للدخل الحقيقي للفرد في سنة ما (سنة المقارنة) إلى سنة أخرى (سنة الاساس). سنة الأساس هي الفترة الزمنية التي نقيس منها التغير في الظاهرة بينما سنة المقارنة هي الفترة الزمنية التي حصل خلالها تغير الظاهرة.

وسوف نتطرق في هذا الفصل لعدة أنواع من الأرقام القياسية ومنها:

الرقم القياسي البسيط Simple Index Number:

الرقم القياسي البسيط هو الرقم المتمثل من نسبة متغير واحد في فترة المقارنة على نفس المتغير في فترة الأساس. سيكون:

$$\frac{\text{سنة المقارنة}}{\text{سنة الاساس}} \times ١٠٠$$

$$I_1 = \frac{p_n}{p_o} \times 100\%$$

حيث أن (p_n) سنة المقارنة و (p_o) سنة الاساس.

مثال (1):

البيانات التالية تمثل أسعار برميل النفط للأعوام التالية:

السنة	2000	2001	2002	2003	2004	2005
السعر بالدولار	30	34	38	42	45	48

المطلوب: حساب الارقام القياسية البسيطة للأعوام (2000، 2004). علماً بأن سنة الأساس هي عام (2002).

الحل:

الرقم القياسي البسيط لعام (2000) تساوي:

$$I_{2000} = \frac{30}{38} \times 100\% = 78.95$$

وبما أن سنة الأساس دائماً هي 100% فإن ذلك يعني أن أسعار عام (2000) هي أقل من عام (2002) (سنة الأساس) بنسبة (21.05%) بمعنى آخر ان أسعار عام (2000) كانت تقريباً ثلاثة أرباع للأسعار عام (2002).

والرقم القياسي البسيط لعام (2004) هو:

$$I2004 = \frac{45}{38} \times 100\% = 118.42$$

أي أن أسعار عام (2004) أكثر من أسعار عام (2002) بـ:

$$118.42 - 100 = 18.42\%$$

وهي نسبة مجموع أسعار عدة سلع في عام عين (سنة الأساس) إلى مجموع أسعار هذه السلع في عام آخر (سنة المقارنة):

$$Ip\ (a) = \frac{\Sigma p_n}{\Sigma p_o} \times 100\%$$

حيث (Ip a) يعني p(price) الأسعار و (a) تدل على التجميعي (aggregate) و (pn) سنة المقارنة و (po) سنة الأساس.

مثال (2):

أحسب الرقم القياسي التجمعي البسيط لبعض المواد الاستهلاكية كما في الجدول التالي:

السلعة	اسعار عام (2000) (كغم) 15(po)	أسعار عام (2005) (كغم) (pn)
السكر	300	400
الحليب	1500	2500
الأرز	250	380
الشاي	1000	1400
المجموع	3050	4080

الحل:

نجمع أسعار عام (2000) وهي سنة الأساس (po) ومن ثم نجمع أسعار عام 2005 (pn) بجمع السلع ومن ثم نطبق المعادلة:

$$Ip(a) = \frac{\sum p_n}{\sum p_o} \times 100\%$$

$$Ip(a)= \frac{4680}{3050} \times 100\%$$

$$Ip(a) = 153.44$$

أي أن أسعار السلع عام (2005) أرتفعت بنسبة 53.44% عن أسعار عام (2000).

الرقم القياسي النسبي البسيط للأسعار:Simple Relative Price Index

وهو الوسط الحسابي للأرقام القياسية للسلع بمعنى آخر نجد الرقم القياسي لكل سلعة ثم نجد الوسط الحسابي لهذه الأرقام القياسية ويكون كما يلي:

$$Ip(r)= \frac{1}{m} \sum \frac{p_n}{p_o} \times 100\%$$

حيث أن (r) تدل على النسبي (m) عدد السلع.

مثال (3):

أحسب الرقم القياسي النسبي البسيط للأسعار في المثال (2-7) السابق باعتبار أن عام (2000) هي سنة الاساس.

الحل:

الرقم القياسي النسبي البسيط لأسعار (2005) هو:

$$Ip(r)\ 2005 = \frac{1}{4}\left[\frac{400}{300}+\frac{2500}{1500}+\frac{380}{250}+\frac{1400}{1000}\right]\times 100\%$$

$$= \frac{1}{4}[1.33+1.66+1.52+1.4]\times 100\%$$

$$= \frac{1}{4}[5.91] \times 100$$

$$= 1.4775 \times 100$$

$$= 147.75\%$$

أي أن اسعار عام (2005) أرتفعت عن أسعار عام (2000) بنسبة (47.75%).

طريقة لاسبير Laspeyrs Method:

ونستعمل هذه الطريقة الكميات المستهلكة والقيمة التقديرية للكميات المستهلكة في سنة الأساس كأوزان لأسعار المواد الداخلة في حساب الرقمين القياسين التجميعي والنسبي..

رقم لاسبير التجميعي للأسعار Laspeyrs Aggregate Price:

وهو نسبة ما يتم إنفاقه في سنة المقارنة (pn) إلى ما يتم إنفاقه في سنة الأساس (po) على جميع السلع إذا أردنا استهلاك نفس كميات سنة الأساس في الحالين وتكون الصيغة كما يلي:

$$Ip(L) = \frac{\sum P_n Q_o}{\sum P_o Q_0} \times 100$$

حيث أن Pn: هي أسعار سنة المقارنة.

Po: هي أسعار سنة الأساس.

Q0: هي الكميات المستهلكة لسنة الأساس.

L: هي لاسبير.

مثال (4):

أحسب رقم لاسبير التجميعي للأسعار لعام (2007) من الجدول التالي والذي يمثل أسعار عدد من السلع فلس/كغم وبإعتبار أن سنة الأساس هي (2001).

السلعة	اسعار عام (2001) (po)	كميات استهلاك (2001) (Qo)	أسعار عام (2007) (Pn)	كميات استهلاك (2007) (Qn)	(Pn.Qo)	(Po.Qo)
السكر	300	10	450	11	4500	3000
الأرز	350	8	420	9	3360	2800
العدس	420	6	640	8	3840	2520
المجموع					11700	8320

الحل:

نضيف عمودين إحدهما (PnQo) أي أسعار سنة المقارنة مضروب بكميات سنة الاساس والآخر أسعار سنة الأساس مضروب بكميات سنة الاساس ثم يأخذ حاصل الجمع لكل عمود ونطبق الصيغة:

$$Ip(L) = \frac{\sum P_n Q_o}{\sum P_o Q_0} \times 100$$

$$Ip(L) = \frac{11700}{8320} \times 100$$

$$Ip(L) = 140.63$$

أي أن أسعار عام (2007) للسلع ارتفعت بنسبة (40.63%).

رقم لاسبير النسبي للأسعار Laspeyrs Percentage Price:

ويكون بالصيغة التالية:

$$Ip(rL) = \sum \left(\frac{Pn}{Po}\right).Wo.100\%$$

حيث أن:

$$Wo = \frac{PoQo}{\sum PoQo}$$

والرمز (rL) يدل على لاسبير النسبي.

مثال (5):

أحسب رقم لاسبير النسبي للأسعار من المثال السابق (4-7).

الحل:

نستعمل عمود (PoQo) ومن ثم نجد (Wo) ومن الممكن عمل عمود له علماً بـأن مجموع (Wo) دائماً يساوي واحد.

Wo	PoQo
$0.36 = \frac{3000}{8320}$	3000
$0.34 = \frac{2800}{8320}$	2800
$0.30 = \frac{2520}{8320}$	2520
1	8320

والآن نحسب رقم لاسبير النسبي للأسعار وهو:

$$Ip(rL) = \left[\sum \frac{Pn}{Po} \times Wo\right] \times 100$$

$$= \left[\frac{450}{300} \times 0.36 + \frac{420}{350} \times 0.34 + \frac{640}{420} \times 0.30\right] \times 100$$

$$=[1.5 \times 0.36 + 1.2 \times 0.34 + 1.52 \times 0.30] \times 100$$

$$= [0.54 + 0.408 + 0.456] \times 100$$

$$= 1.404 \times 100$$

$$= 140.4$$

أي أن رقـم لاسـبير النسـبي للأسـعار (2007) ارتفـع بنسـبة (40.4%) عـن أسـعار (2001).

طريقة باش Passche Method:

وتسـتعمل هـذه الطريقـة الكميـات المسـتهلكة وأسـعارها في سـنة المقارنـة واوزان لأسعار المواد الداخلة في حساب الرقمين التجميعي والنسبي.

رقم باش التجميعي للأسعار Passche Aggregate Price:

وهو نسبة ما يتم إنفاقه في سنة المقارنة إلى ما تم انفاقه في سنة الأساس على جميع السلع في حالة استهلاكنا لنفس كميات سنة المقارنة وتكون الصيغة كما يلي:

$$Ip(p) = \frac{\sum P_n Q_o}{\sum P_o Q_0} \times 100$$

حيث أن Pn: هي أسعار سنة المقارنة.

Po: هي أسعار سنة الأساس.

Qn: هي الكميات المستهلكة (المبيعة) في سنة المقارنة.

P: يدل على باش

مثال (6):

أحسب رقم باش التجميعي للأسعار للمثال السابق (Qn) (Pn) (Qo).

السلعة	اسعار عام (2001) (po)	كميات استهلاك (2001) (Qo)	أسعار عام (2007) (Pn)	كميات استهلاك (2007) (Qn)	(Pn.Qo)	(Po.Qo)
السكر	300	10	450	11	3300	4950
الأرز	350	8	420	9	3150	3780
العدس	420	6	640	8	3360	5120
المجموع					9810	13850

الحل:

نضيف عمودين إحدهما (PnQo) أي حاصل ضرب أسعار سنة المقارنة مضروب بكميات سنة المقارنة والعامود الآخر أسعار سنة الأساس مضروب بكميات سنة المقارنة ثم يأخذ حاصل الجمع هما ونطبق الصيغة:

$$Ip(p) = \frac{\sum P_n Q_o}{\sum P_o Q_0} \times 100$$

$$Ip(p) = \frac{13850}{9810} \times 100 = 141.18$$

أي أن رقم باش التجميعي للأسعار لسنة (2007) أرتفع بنسبة (41.18%) عـن سـنة (2001).

- رقم باش النسبي للأسعار:

$$Ip(rp) = \sum \frac{p_n}{p_o} \times Wn \times 100$$

حيث أن (rp) تدل على باش النسبي.

$$Wn = \frac{p_n Q_n}{\sum p_n Q_n}$$

مثال (7):

أحسب رقم باش النسبي للاسعار حسب المعطيات في المثال السابق (7).

الحل:

نستخدم معطيات المثال السابق وبشكل خاص عمود (Pn.Qn) ومن ثم نجد (Wn).

Wn	PnQn
$0.36 = \frac{4950}{13850}$	44950
$0.27 = \frac{3780}{13850}$	3780
$0.37 = \frac{5120}{13850}$	5120
1	13850

ويكون رقم باش النسبي للأسعار كما يلي:

$$Ip(rp) = \left[\frac{450}{300} \times 0.36 + \frac{420}{350} \times 0.27 + \frac{640}{420} \times 0.37\right] \times 100$$

$$=[1.5 \times 0.36 + 1.2 \times 0.27 + 1.52 \times 0.37] \times 100$$

$$= [0.54 + 0.324 + 0.5624] \times 100$$

$$= 1.4264 \times 100$$

$$= 142.64$$

أي أن رقم باش النسبي للأسـعار (2007) زادت بمقـدار (42.64%) عـن أسـعار عـام (2001).

طريقة فشر Fisher's Method:

وهي عبارة عن دمج طريقة لاسبير مع طريقة باش وتحت الجذر.

رقم فيشر التجميعي للأسعار هو:

$$Ip(f) = \sqrt{Ip(L).Ip(p)}$$

رقم فيشر النسبي للاسعار هو:

$$Ip(rf) = \sqrt{Ip(rL).Ip(rp)}$$

رقم باش التجميعي للأسعار Ip(p) هو (141.18).

ورقم لاسبير التجميعي للاسعار (140.63).

رقم باش النسبي للاسعار Ip(rp) هو (142.64)

ورقم لاسبير النسبي للاسعار Ip(rL) هو (140.4)

وحسب هذه المعطيات السابقة يكون رقم فيشر التجميعي للاسعار هو:

$$Ip(f) = \sqrt{140.63 \times 141.18}$$

$$= \sqrt{19854.1434} = 140.9$$

ورقم فيشر النسبي للاسعار هو:

$$Ip(rf) = \sqrt{140.4 \times 142.64}$$

$$= \sqrt{20026.656} = 141.52$$

تمارين الفصل التاسع

1. وضح المقصود بالارقام القياسية.
2. الجدول التالي يمثل أسعار اللحوم لخمسة أعوام وهي كما يلي:

السنة	1998	2002	2003	2004	2005
السعر بالدولار	3	4	5	5.5	7

المطلوب: حساب الرقم القياسي البسيط للاعوام (1998، 2005) و (2002، 2004).

3. يمثل الجدول التالي اسعار وكميات لمجموعة من السلع لمدة عامين وهي كما يلي:

السلعة	أسعار عام (1999)	كميات عام (1999)	أسعار عام (200)	كميات عام (2006)
الحليب	1600	18	2600	22
الشاي	900	10	1400	15
الطحين	200	20	250	30
الارز	400	5	600	8

المطلوب:

أ-حساب الرقم القياسي التجميعي البسيط.

ب- الرقم القياسي النسبي البسيط.

ج- رقم لاسبير التجمعي للأسعار.

د— رقم لاسبير النسبي للأسعار.

هـ- رقم باش التجمعي للأسعار.

و- رقم باش النسبي للاسعار.

4. الجدول التالي يمثل أسعار وكميات لمجموعة من السلع خلال فترتين:

السلعة	أسعار عام (2000)	كميات عام (2000)	أسعار عام (2008)	كميات عام (2008)
الدجاج	120	100	160	140
السكر	300	10	400	15
الكاز	100	30	400	35

المطلوب:

أ-حساب رقم فيشر التجمعي للأسعار.

ب-حساب رقم فيشر النسبي للأسعار.

تمارين عامة

السؤال الأول: يتكون من (30) فرعاً كل فرع له أربع خيارات إحداها صحيح حدد.

1-1يهدف علم الاحصاء إلى الوصول لـ:

أ- تحليل البيانات   ب- عرض البيانات

ج- تبويب البيانات   د- قرارات ملائمة.

1-2 يسمى الإحصاء الاستنتاجي بـ:

أ- الاحصاء الوصفي   ب- الإحصاء العلمي

ج- الإحصاء الاستدلالي   د- الإحصاء القياسي

1-3 من أنواع العينات، عينات غير احتمالية ومنها:

أ- العينة المنتظمة   ب- المعاينة بالاختيار السهل

ج- العينة العنقودبة   د- العينة العشوائية البسيطة

1-4 إذا اردنا عينة مكونة من (40) موظف من مجتمع مصنع المكون من (800) عامل عادي و (400) عامل فني فكم تكون العينة من كل فئة تقريباً.

أ- 27، 13   ب- 26، 14

ج- 30، 10   د- 25، 15

1-13هناك مجموعتان الأولى حجمها 30 ووسطها الحسابي 45 والمجموعة الثانية حجمها 60 ووسطها الحسابي 75 فإن الوسط الحسابي المرجح للمجموعتين معاً هو:

أ- 48   ب- 49

ج- 63   د- 65

هـ- غير ذلك

1-14 (6، 9، 7، 14) هذه الارقام لدينا ثم قمنا بطرح واحد من كل منها فإن الوسط الحسابي الجديد يكون:

أ- 6   ب- 7

ج- 8   د- 9

1-15 من خصائص الوسيط:

أ- يمكن إيجاده بيانياً   ب- يتأثر بالقيم المتطرفة.

ج- لا يتأثر بعدد القيم والمشاهدات   د-لا يمكن حسابه من الجداول المفتوحة

1-16 إذا كانت أكبر مشاهدة هي (182) ومدى التوزيع يساوي (112) فإن اصغر مشاهدة هي:

أ- 35   ب- 70

ج- 90   د- 147

17-1 إذا كان المئين 75 هو (90) والمئين 50 هو (70) والمئين 25 هو (40) في هذه الحالة نصف المدى الربعي يكون:

أ- 20 ب- 25

ج- 50 د- غير ذلك.

18-1إذا كان الوسط الحسابي لمجموعة من البيانات هو (9) والوسيط (11) أثر على هذه البيانات التمويل الخطي التالي ($y=3\times-6$) فإن الوسط الحسابي بعد التعديل يكون:

أ- 20 ب- 21

ج- 27 د- غير ذلك

- استخدم هذه البيانات للاجابة على سؤال 1-19 حى 1-22

الفئات	15-10	21-16	27-22	33-28	39-34	45-40
التكرار	10	9	8	7	6	10

19-1 الوسط الحسابي حسب الجدول اعلاه هو:

أ- 26.28 ب- 26.82

ج- 28.28 د- غير ذلك

20-1 المنوال هو:

أ- 12.5، 42.5 ب- 12.5

ج- 42.5 د- الفئة الأولى والأخيرة

ضع علامة أمام الإجابة الصحيحة لكل سؤال من الأسئلة التالية:

س 1 / تنقسم البيانات الى: بيانات وصفية وبيانات كمية.

أ: صح ب: خطأ

س 2 / المدى = اكبر قيمة + اصغر قيمة.

أ: صح ب: خطأ

س 3 / يمكن أن يكون للبيانات أكثر من وسط حسابي.

أ: صح ب: خطأ

س 4 / يمكن أن يكون للبيانات أكثر من منوال.

أ: صح ب: خطأ

س -5 / أحيانا لا نجد منوال لبعض البيانات.

أ: صح ب: خطأ

س -6 / يمكن إيجاد الوسط الحسابي من الجداول التكرارية.

أ: صح ب: خطأ

س 7 / يمكن إيجاد الوسيط من المنحني المتجمع الصاعد.

أ: صح ب: خطأ

س 8 / مركز الفئة = (الحد الأعلى للفئة + الحد الأدنى للفئة) ÷ 2

أ: صح ب: خطأ

س 9 / مركز الفئة = (الحد الأعلى للفئة - الحد الأدنى للفئة) ÷ 2

أ: صح ب: خطأ

س 10 / يستخدم المنحني المتجمع الصاعد في إيجاد:

أ: الوسط الحسابي. ب: الوسيط. ج: المنوال

س 11 / البيانات التالية تمثل أعمار عينة من الموظفين:

22 , 25 , 34 , 36 , 20 , 27 , 33 ما هي قيمة الوسيط:......

أ: 33 ب: 27 ج: 25

س 39 / البيانات التالية تمثل أعمار عينة من الموظفين:

22 , 25 , 34 , 36 , 20 , 27 , 33, 21 , 30 , 29 ما هي قيمة الوسيط:....

أ: 27 ب: 28 ج: 29

س 40 / البيانات التالية تمثل أعمار عينة من الموظفين: 22 , 25 , 34 , 35 , 20 , 27 , 33 ما هي قيمة الوسط الحسابي:......

أ: 33 ب: 30 ج: 28

س 41/ البيانات التالية تمثل أعمار عينة من الموظفين: 21 , 25 , 34 , 36 , 25 , 27 , 33, 25 , 34 , 25 ما هي قيمة المنوال:....

أ: 34 ب: 30 ج: 25

س 42 / البيانات التالية تمثل أعمار عينة من الموظفين: , 20 , 31 , 36 , 22 , 27 , 33, 26 , 34 , 25 ما هي قيمة المنوال:.......

أ: لا يوجد منوال ب: 30 ج: 28

س60 / الجدول التكراري التالي يبين توزيع الأوزان لعينة من الطلاب.

فئات الوزن	60-	64-	68-	72-	76-	80-	84-88	$\sum$
العدد f	5	12	20	26	20	12	5	100

الوسط الحسابي =..............

أ: 77 ب: 74 ج: 80 د: 79

س61 / الجدول التكراري التالي يبين توزيع الأوزان لعينة من الطلاب.

فئات الوزن	60-	64-	68-	72-	76-	80-	84-88	$\sum$
العدد f	5	12	20	26	20	12	5	100

الوسيط =...............

أ: 74 ب: 76 ج: 78 د: 80

س62 / الجدول التكراري التالي يبين توزيع الأوزان لعينة من الطلاب.

فئات الوزن	60-	64-	68-	72-	76-	80-	84-88
عدد الطلبة	5	12	20	26	20	12	5

المنوال يساوي:...............

أ: 68 ب: 70 ج: 72 د: 74

س 64 / الجدول التالي يمثل توزيع الدرجات في احد الاختبارات لمجموعة من الطلاب:

الدرجات	30-	40-	50-	60-	70-	80-	90-100
عدد الطلبة	2	5	8	15	8	5	2

المنوال =.........................

أ: 65 ب: 77 ج: 60 د: 75

س 65 / الجدول التالي يمثل توزيع الدرجات في احد الاختبارات لمجموعة من الطلاب:

الدرجات	30-	40-	50-	60-	70-	80-	90-100
عدد الطلبة	2	5	8	15	8	5	2

الوسط الحسابي =.........................

أ: 61 ب: 64 ج: 65 د: 69

س 67 / الجدول التالي يمثل توزيع الدرجات في احد الاختبارات لمجموعة من الطلاب:

الدرجات	30-	40-	50-	60-	70-	80-	90-100
عدد الطلبة	2	5	8	15	8	5	2

الوسيط يساوي:...................

أ: 67.5 ب: 65 ج: 66 د: 69

س 69 / الجدول التالي يبين توزيع الأعمار لعينة من الطلاب في كلية العلوم الانسانية:

الأعمار	16-	18-	20-	22-	24-26	$\sum$
عدد الطلبة	4	10	18	12	6	50

المنوال =.................

أ: 21.09 ب: 26.09 ج: 23.09

المراجع

المراجع العربية:

1-أحمد قاسم وآخرون، المدخل على علم الاحصاء، منشورات جامعة حلب، 1988.

2-محمود المشهداني وآخرون، الاحصاء، دار الحكمة للنشر والتوزيع، بغداد، 1989.

3-عبد الرحمن عدس، مبادئ الاحصاء في التربية وعلم النفس، دار الفكر للطباعة والنشر، عمان – الاردن، 1995م.

4-أحمد سرحان، مقدمة في طرق التحليل الإحصائي، معهد البحوث والدراسات الإحصائية، القاهرة، 1974.

المراجع الاجنبية:

1. Hald, Anders (1998) "A History of Mathematical Statistics (1970-1930) Wiley Series in probability and Statistics.
2. Porter, T. M. (1986) "The Rise of Statistical Thinking (1820-1900), Princton University Press.
3. Gani, J (Editor) (1982) "The Making of Statisticians" , Springer – Verlag.
4. Pearson , E.S. and Kendal, Sir M. (editors) (1978) "Studies in the History of Statistics and Probability ", vol, I, Charles Griffin, London.
5. Kendal, Sir M. & Plackett, R L (Editors) (1977) "in the History of Statistics and probability "Vol. II, Charles Giffin, London.
6. Dixon, W. J. nd Mood, A. M. (1948) "A method for Obtaining and Analysing Sensitivity Data" , Vol. 43, pp. 109-126.
7. Adrich , J. (2005): figures from the History of Probability and Statistics "Uinversity of Southampton, Southamton, UK.

8. Moskowite,H, wright, G. “Statistics for management and Economics”. Columbus , Charles E. Merril publishing company, 1985.
9. Lee, c. “Statistics for Business and Finacial Economics”. Lexington, D.C. Health and company , 1993.

الفهرس